

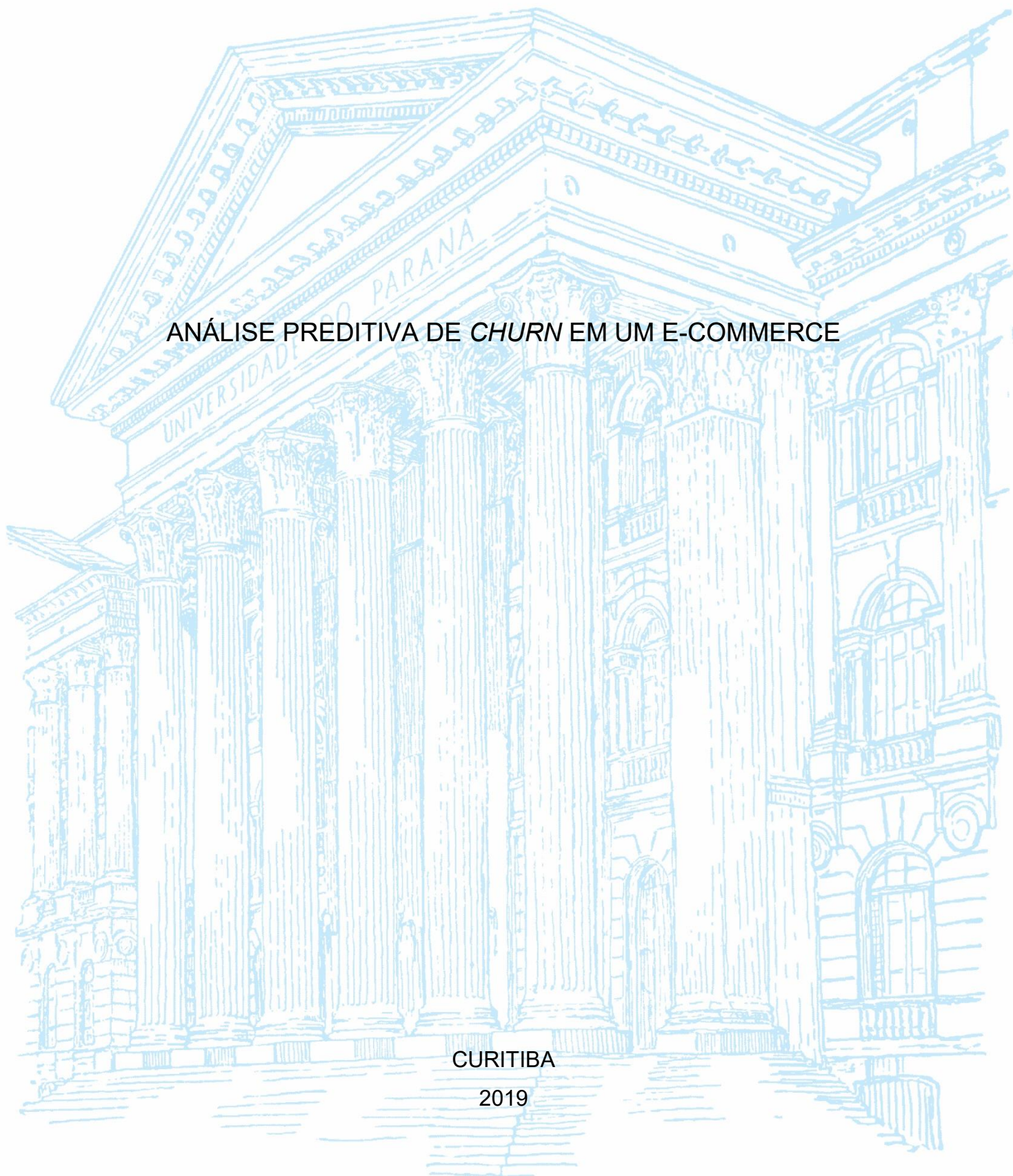
UNIVERSIDADE FEDERAL DO PARANÁ

BRUNA FRANCIANY GIRATA LUIZON

ANÁLISE PREDITIVA DE *CHURN* EM UM E-COMMERCE

CURITIBA

2019



BRUNA FRANCIANY GIRATA LUIZON

ANÁLISE PREDITIVA DE *CHURN* EM UM E-COMMERCE

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção de grau de Bacharel no Curso de Gestão da Informação, Departamento de Ciência e Gestão da Informação do Setor de Ciências Sociais Aplicadas, da Universidade Federal do Paraná.

Orientadora: Prof^a. Dr^a. Denise Fukumi Tsunoda.

CURITIBA

2019

AGRADECIMENTOS

Um ciclo da minha vida se encerra para que outro possa iniciar. Novos desafios, novas metas, novos objetivos e uma nova "Bruna" está surgindo e não posso deixar de agradecer a quem tanto me auxiliou e me deu suporte todo esse período.

Agradeço primeiramente a Deus por toda força, ânimo e coragem que me ofereceu para alcançar minha meta.

Aos professores que com toda paciência e esforço transmitiu um pouco de seus conhecimentos. Foram eles que me deram recursos e ferramentas para evoluir um pouco mais todos os dias e sair do meu "feitiço do tempo".

Um agradecimento especial para a professora Dr^a Denise Tsunoda que com toda paciência me orientou, ensinou e me deu forças principalmente no período final da faculdade.

A você Pai, Moacir. Obrigada pelas resilientes madrugadas me ouvindo reclamar e ainda assim não me deixando desistir. Obrigada por me dizer que tudo ia ficar bem e por demonstrar todo orgulho que sente de mim! Agradeço a Carla, obrigada por ser além de minha irmã, minha amiga que por todo esse período me apoiou e deu os melhores conselhos, me motivando a ser melhor cada vez mais.

Mayla, amiga de longa data, agradeço por estar presente em todos os momentos que precisei! Fernando, amigo que conheci ao longo dessa jornada, obrigada pelo companheirismo e compreensão.

A todas as pessoas que de uma alguma forma me ajudaram a acreditar em mim eu quero deixar um agradecimento!

RESUMO

Estudo sobre a aplicação de *machine learning* em análises preditivas de *churn* em e-commerce. Busca entender o conceito de clientes *churn* dentro de organizações onde não há assinatura de um serviço e/ou produto, bem como o impacto da tecnologia na evolução do marketing e das novas formas de comércio. Analisa documentos recuperados sobre o tema na base de dados Scopus, fazendo uma revisão sistemática e análise bibliométrica dos documentos levantados. Há uma série de técnicas de mineração de dados aplicadas para a previsão de *churn*, tais como redes neurais artificiais, árvores de decisão e máquinas vetoriais de suporte. Apresenta a relação entre *machine learning* e a predição de *churn*. Conclui-se na primeira etapa a recência do assunto e a falta de estudos aprofundados sobre o tema, além de que os algoritmos mais utilizados envolvem árvore de decisão, Support Vector Machine, Rede Neural Artificial, *Random Forest* e Regressão Logística. Comenta que a maioria dos estudos existentes relacionados com a previsão do *churn* de clientes são análises estáticas e não estão bem adequados à realização de monitorizações individuais e dinâmicas, já que a análise de dados estáticas não fornecem monitoramento dinâmico do *churn* do cliente. No segundo momento do estudo, foram selecionados três métodos e aplicados em uma base de um e-commerce na área de varejo. Os métodos escolhidos foram: Árvore de Decisão (J48), Rede Neural Artificial (*Multilayer Perceptron*) e *Support Vector Machine* (SMO). Concluiu que os três modelos apresentaram resultados similares quando analisados a taxa de acertos do modelo, porém o que apresenta melhor tempo de execução do modelo é a Árvore de Decisão J48.

Palavras-chave: Previsão do *churn* do cliente. Técnicas de previsão. Previsão. *Machine learning Churn* em e-commerces. Comércio electrónico.

ABSTRACT

Study on the application of machine learning in predictive analysis of churn in e-commerce. It seeks to understand the concept of churn customers within organizations where there is no subscription to a service and/or product, as well as the impact of technology on the evolution of marketing and new forms of commerce. It analyzes documents recovered on the theme in the Scopus database, making a systematic review and bibliometric analysis of the documents raised. There are a number of data mining techniques applied to churn prediction, such as artificial neural networks, decision trees and supporting vector machines. It presents the relationship between machine learning and churn prediction. The first stage concludes with the recency of the subject and the lack of in-depth studies on the subject, in addition to the fact that the most used algorithms involve decision tree, Support Vector Machine, Artificial Neural Network, Random Forest and Logistic Regression. He comments that most of the existing studies related to customer churn forecasting are static analyses and are not well suited to perform individual and dynamic monitoring, since static data analysis does not provide dynamic monitoring of customer churn. In the second phase of the study, three methods were selected and applied on an e-commerce basis in the retail area. The methods chosen were: Decision Tree (J48), Artificial Neural Network (Multilayer Perceptron) and Support Vector Machine (SMO). He concluded that the three models presented similar results when analyzing the hit rate of the model, but the one that presents the best execution time of the model is the Decision Tree J48.

Keywords: Customer churn prediction. Prediction techniques. Prediction. Machine learning Churn in e-commerces. Electronic commerce.

LISTA DE FIGURAS

FIGURA 1: PUBLICAÇÕES POR ANO	14
FIGURA 2: PAÍSES CORRESPONDENTE DOS AUTORES.....	14
FIGURA 3: A ROTA GLOBAL DO E-COMMERCE.....	17
FIGURA 4: LINHA DO TEMPO DA INFORMAÇÃO.....	18
FIGURA 5: ETAPAS DO PROCESSO DE KNOWLEDGE DISCOVERY IN DATABASES (KDD).....	21
FIGURA 6: CLASSIFICAÇÃO DAS ESTRATÉGIAS DE APRENDIZADO DE MÁQUINA.....	21
FIGURA 7: RESUMO DAS ETAPAS DO PROCESSO DE UM MODELO PREDITIVO.....	26
FIGURA 8: CARACTERIZAÇÃO DA PESQUISA.....	31
FIGURA 9: ETAPAS DA REALIZAÇÃO DA PESQUISA.....	32
FIGURA 10: METODOLOGIA PRISMA ADAPTADA.....	35
FIGURA 11: FLUXOGRAMA DO ALGORITMO DE ÁRVORE DE DECISÃO	38
FIGURA 12: CLASSIFICAÇÃO ESTRUTURAL E FUNCIONAL DE REDES NEURAIS ARTIFICIAIS	39
FIGURA 13: RELAÇÃO DE NÚMERO DE PUBLICAÇÕES COM LINHA DE TENDÊNCIA.....	42
FIGURA 14: PAÍSES ORIGEM DAS PUBLICAÇÕES.....	43
FIGURA 15: NUVEM DE PALAVRAS: TÍTULOS E PALAVRAS-CHAVES PLUS	44
FIGURA 16: PRIMEIRA LEI DE ZIPF.....	45
FIGURA 17: PALAVRAS-CHAVE DOS AUTORES E RESUMO.....	46
FIGURA 18: MÉTODOS UTILIZADOS	49
FIGURA 19: SÍNTESE DA METODOLOGIA UTILIZADA.....	50
FIGURA 20: INFORMAÇÕES PROCESSADAS.....	51
FIGURA 21: REPRESENTAÇÃO GRÁFICA DOS ATRIBUTOS GERADA PELO WEKA	53
FIGURA 22: PARÂMETROS DEFAULT J48	54
FIGURA 23: RESULTADOS DO J48 NO WEKA	56
FIGURA 24: MÉTRICAS DE DESEMPENHO POR CLASSE AVALIADA	56
FIGURA 25: MATRIZ DE CONFUSÃO.....	57

FIGURA 27: REPRESENTAÇÃO GRÁFICA DA ÁRVORE J48.....	59
FIGURA 28: PARÂMETROS DEFAULT <i>MULTILAYER PERCEPTRON</i>	60
FIGURA 29: RESULTADOS DO <i>MULTILAYER PERCEPTRON</i>	63
FIGURA 30: PARÂMETROS DEFAULT SMO	64
FIGURA 31: RESULTADOS DO SMO	65

LISTA DE TABELAS

TABELA 1: PRINCIPAIS ÁREAS DE PESQUISA	13
TABELA 2: DESCRIÇÃO DA BASE DE DADOS	36
TABELA 3: LEI DE BRADFORD.....	43
TABELA 4: FREQUÊNCIA DAS PALAVRAS-CHAVE PLUS.....	44
TABELA 5: FREQUÊNCIA DAS PALAVRAS-CHAVE PARA APLICAÇÃO DA LEI DE ZIPF	45
TABELA 6: ARTIGOS MAIS CITADOS	46
TABELA 7: DADOS DISCRETIZADOS.....	52
TABELA 8: DESCRIÇÃO DA BASE DE DADOS APÓS DISCRETIZAÇÃO.....	52
TABELA 9: COMPARATIVO DOS RESULTADOS OBTIDOS DOS MODELOS.	66
TABELA 10: MÉDIA DE ACURÁCIA DOS ALGORITMOS.....	68

SUMÁRIO

1 INTRODUÇÃO	10
1.1 PROBLEMATIZAÇÃO	11
1.2 OBJETIVOS	12
1.2.1 OBJETIVO GERAL	12
1.2.2 OBJETIVOS ESPECÍFICOS	13
1.3 JUSTIFICATIVA	13
2 REVISÃO DA LITERATURA	16
2.1 E-COMMERCE	16
2.2 MARKETING DIGITAL	17
2.3 CUSTOMER <i>CHURN</i>	20
2.4 MINERAÇÃO DE DADOS	21
2.5 MACHINE LEARNING	23
2.6 ANÁLISE PREDITIVA	25
2.7 RELAÇÃO ENTRE MACHINE LEARNING E <i>CHURN</i> PREDICTION	27
3 METODOLOGIA	30
3.1 CARACTERIZAÇÃO DA PESQUISA	30
3.2 ETAPAS, MATERIAIS E MÉTODOS	31
3.3 METODOLOGIA PRISMA	34
3.4 BASE DE DADOS E APLICAÇÃO	35
3.5 MÉTODOS DE MACHINE LEARNING	36
3.5.1 ÁRVORE DE DECISÃO (J48)	37
3.5.2 REDE NEURAL ARTIFICIAL (RNA)	38
3.5.3 MÁQUINA DE VETORES DE SUPORTE (SVM)	40
4. ANÁLISE E DISCUSSÃO DOS DOCUMENTOS RECUPERADOS	42
4.1 RESULTADOS DAS ANÁLISES QUANTITATIVAS	42
4.2 RESULTADOS DAS ANÁLISES DE CONTEÚDO	46
5. APLICAÇÃO DOS MODELOS DE MACHINE LEARNING	51
5.1.1 ÁRVORE DE DECISÃO (J48)	53
5.1.2 REDE NEURAL ARTIFICIAL (RNA)	59
5.1.3 MÁQUINA DE VETORES DE SUPORTE (SVM)	63
5.2 COMPARAÇÃO DOS RESULTADOS DOS MODELOS APLICADOS	66
6. CONSIDERAÇÕES FINAIS	68

REFERÊNCIAS	71
APÊNDICE A – LISTA DE ARTIGOS.....	77
APÊNDICE B - RESULTADO DO ALGORITMO J48	80
APÊNDICE C - RESULTADO DO SMO	83
APÊNDICE D – RESULTADO DO MULTILAYER PERCEPTRON.....	85

1 INTRODUÇÃO

Com a inserção de novas tecnologias e o ambiente digital nas últimas décadas, o comportamento do consumidor tem sofrido alterações, já que as pessoas estão migrando para a esfera *online*. O consumidor está cada vez mais consciente e utiliza muito da *internet* para fazer pesquisas de preços e comparar as lojas.

A *internet* criou uma nova forma de comercialização, o comércio eletrônico (*e-commerce*), que é a venda de produtos em uma loja virtual. É um mercado em expansão que no Brasil cresceu 12% de faturamento em 2018 e a expectativa é de que cresça cerca de 15% em 2019 (*E-COMMERCE BRASIL*, 2018).

No Brasil as três categorias de e-commerce mais comuns são: *business to consumer* (B2C), que é relação de venda entre empresa (fabricantes, revendedoras ou varejistas) e o consumidor; *business to business* (B2B), vendas exclusivas para empresas e; *consumer to consumer* (C2C), onde qualquer pessoa pode cadastrar um produto e vender. Outros tipos de e-commerce são: *Customer to Business* (C2B) que é a venda de pessoas físicas para pessoas jurídicas; *Business to government* (B2G) ou *Business to Administration* (B2A) é uma negociação entre empresa e o governo, há diversas leis e regras que regem essa negociação; *Citizen to Government* (C2G) negociações entre cidadãos e a administração pública dependem mais Nde medidas de governo eletrônico (e-gov) que permitam que pessoas físicas proponham soluções que contribuam para melhorar a eficiência das ações do Estado. Há ainda os tipos de *e-commerce* que utilizam plataformas específicas o comércio, o *Mobile commerce* (M-Commerce), onde a transação é originada de um dispositivo móvel; *Social commerce* (S-Commerce), utiliza as redes sociais para transações de compra; Modalidade de comércio televisivo (*T-Commerce*) que utiliza a TV Digital para a realizar a venda de produtos que estejam sendo apresentados em algum programa.

Com isso o marketing ganhou uma nova perspectiva, o *marketing* digital, que consiste em um processo destinado a ajudar vender produtos e serviços para um público-alvo usuário de *internet*, junto com ferramentas e serviços *online*, de forma estratégica e coerente com o plano de *marketing* da empresa.

Dentro da área de *marketing* digital há algumas métricas que podem auxiliar os resultados da organização trazendo maior eficácia para a tomada de decisão estratégica. Uma delas é a taxa de *churn*, ou taxa de desistência de clientes, que no ambiente de um comércio eletrônico B2C é o percentual de clientes que deixaram de efetuar a compra.

Mais importante ainda do que a métrica de taxa de *churn*, é evitá-lo, e para isso o *marketing* digital pode se apoiar em técnicas de *machine learning*, já que o marketing tem se tornado uma área cada vez mais precisa e dependente dos dados para provar seus resultados. Ao entender melhor as razões da desistência de compra de um cliente, a empresa possui condições mais adequadas para realizar os ajustes necessários na operação e pode desenvolver um plano de ação mais eficiente.

Machine learning utiliza algoritmos para identificação de padrões de dados e cria conexões entre eles para aprender a executar uma tarefa sem a ajuda humana e de forma inteligente. Esses algoritmos usam análises estatísticas para prever respostas mais precisamente e entregam o melhor resultado preditivo com menos chance de erro.

Com isso, essa pesquisa visa identificar os principais métodos de *machine learning* utilizados para a previsão de *churn* em e-commerces e testá-los utilizando uma base de dados real, buscando identificar o melhor modelo.

1.1 PROBLEMATIZAÇÃO

O consumidor nas últimas décadas anos tem ganhado grande relevância para as empresas, que tentam oferecer o melhor produto para o cliente, com intuito de torná-los fiéis e entender seu comportamento é complexo, já que “Os clientes estão exigindo cada vez mais qualidade e serviço superiores, além de alguma customização. Eles percebem menos diferenças reais entre produtos e mostram menos fidelidade a marcas” (KOTLER, 2000, p.48). O consumidor está cada vez mais consciente e utiliza muito da *internet* para a tomada de decisão de compra. Diferentemente das lojas de varejo tradicionais, as lojas *online* não têm uma interface presencial com o cliente o consumidor não tem contato físico com o

produto, não podendo tocá-lo ou testá-lo, devendo assim confiar no que está sendo apresentado. A forma como a marca se apresenta aos clientes deve ser diferente.

A tecnologia disponível nos dias atuais garante o armazenamento de informações sobre o consumidor, diversos bancos de dados possuem essa função nas empresas, porém a análise desses dados acaba sendo superficial. Os avanços tecnológicos vêm auxiliando as empresas a desenvolver e trabalhar o *marketing* de uma forma mais eficiente, utilizando informações para o seu próprio benefício. O processo de busca por padrões pode ser automatizado por meio de técnicas e ferramentas. “A utilização adequada da mineração de dados permite que as empresas transformem o novo conhecimento em ação e valor de negócios” (KUMAR e REINARTZ, 2012, p.235).

Por essa razão, esta pesquisa visa a aplicação de modelos preditivos de *machine learning* buscando responder a seguinte questão: **Quais modelos preditivos estão sendo utilizados para prever o *churn* de clientes em e-commerce?**

1.2 OBJETIVOS

Os objetivos constituem a finalidade do trabalho e são distinguidos em objetivos gerais e os objetivos específicos. Enquanto os objetivos gerais são aqueles mais amplos, ou seja, são as metas de longo alcance do trabalho, os objetivos específicos são metas mais específicas dentro do trabalho que quando somadas, conduzirão ao desfecho do objetivo geral.

1.2.1 OBJETIVO GERAL

O objetivo geral da pesquisa consiste em uma pesquisa bibliográfica e análise de métodos com o intuito de identificar os modelos utilizados para prever um cliente *churn* em um *e-commerce* utilizando análises preditivas e modelos de *machine learning*.

1.2.2 OBJETIVOS ESPECÍFICOS

Para atingimento do objetivo geral são definidos alguns objetivos específicos, sendo eles:

- pesquisar na literatura as ferramentas e técnicas que são mais utilizadas para a aplicação de modelo preditivo de *churn*;
- definir métodos de *machine learning* para predição de desistência de compra;
- analisar as principais ferramentas e técnicas visando sua aplicabilidade no modelo de negócio de um *e-commerce*.

1.3 JUSTIFICATIVA

Foi realizado um levantamento em 14 de maio de 2019 na base de dados Scopus, utilizando os termos "*machine learning*" AND "*E-commerce*" AND "*Churn prediction*", excluíram-se livros, revistas e publicações comerciais pôr em sua maioria não disponibilizar o acesso ao texto completo, consideram-se ainda todos os idiomas e não houve recorte de período. Foram obtidos 131 resultados. As principais áreas de pesquisas são apresentadas na tabela 1.

TABELA 1 - PRINCIPAIS ÁREAS DE PESQUISA

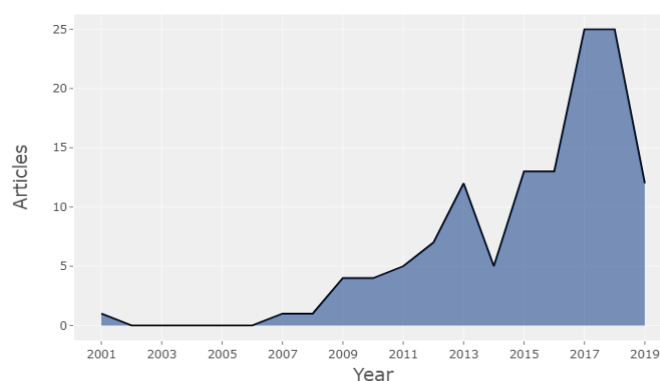
Tema de pesquisa	Ocorrências	Grupo
data mining	39	Mineração de dados
churn prediction	18	Mineração de dados
customer relationship management	17	Mineração de dados
machine learning	15	Mineração de dados
e-commerce	12	Comércio eletrônico
customer churn	12	Rotatividade de clientes
classification	7	Mineração de dados
support vector machine	6	Comércio eletrônico
predictive analytics	5	Análise preditiva

FONTE: AUTORA (2019)

Nota-se que os principais temas de pesquisa estão relacionados com a área de mineração de dados. A primeira publicação sobre o tema foi em 2001, sendo a

próxima somente em 2007. É um tema que possui maior relevância nos últimos quatro anos, cerca de 68,22% das publicações acontecem após 2015, conforme apresentado na figura 1.

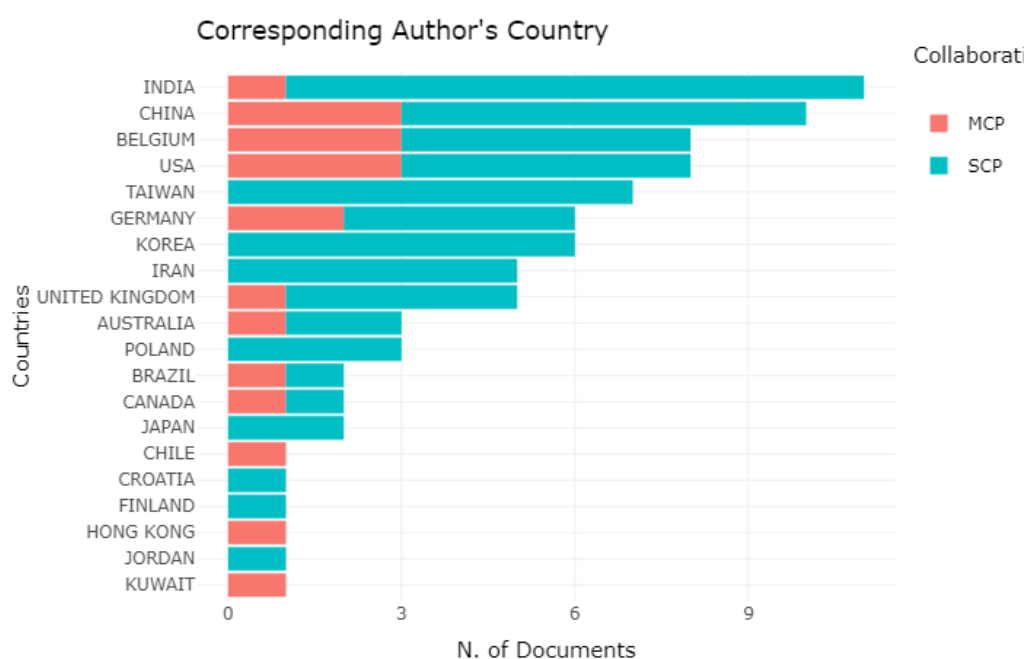
FIGURA 1: PUBLICAÇÕES POR ANO



FONTE: BIBLIOSHINY FOR BIBLIOMETRIX (ARIA)

Os principais países dos autores da coleção são Índia, China e Bélgica. Ordenando os 32 países por quantidade de autores, que tiveram publicações, o Brasil está na 12ª colocação, com apenas dois documentos publicados, conforme apresentado na figura 2.

FIGURA 2: PAÍSES CORRESPONDENTE DOS AUTORES



FONTE: BIBLIOSHINY FOR BIBLIOMETRIX (ARIA)

Dos artigos publicados por brasileiros, o primeiro “*Fraud analysis and prevention in e-commerce transactions*” de Evandro Caldeira, Gabriel Brandão e Adriano Pereira (2014), aplica e avalia técnicas de mineração de dados e aprendizado de máquina para identificar fraudes em transações eletrônicas, em operações de cartão de crédito realizadas por *gateways* de pagamento via *Web*, utilizando um conjunto de dados reais do serviço de pagamento eletrônico brasileiro. O segundo “*Support Vector Clustering for Customer Segmentation on Mobile TV Service*”, por Pedro Albuquerque, Solange Alfinito e Claudio Torres (2015), apresenta uma proposta de segmentação de clientes através do *Support Vector Clustering* (SVC), uma técnica que vem ganhando destaque na literatura acadêmica devido aos bons resultados obtidos, aplicado em uma amostra de consumidores do serviço de TV no Brasil. Ambos não estão diretamente relacionados com o tema principal desta pesquisa. Foram excluídos 107 documentos, após a leitura dos títulos e resumos, por se tratarem de análise preditiva em empresas de telecomunicações, clusterização de cliente para classificação de vendas, *churn* de clientes que já possuíam vínculo com a empresa, *machine learning* aplicado em logística, previsão de fraudes e análises de *churn* para B2B. Mantiveram-se 21 artigos (anexo A) que estão relacionados com o tema.

Além da falta de estudos no Brasil na área pesquisada, têm-se motivação acadêmica e pessoal com relação ao tema. Durante o período da faculdade foi possível o contato com a área de *e-commerce* que incluiu a vivência do cotidiano de uma loja virtual de grande porte tendo acesso à todo e qualquer dado, acarretando em um interesse pela antecipação de tomadas de decisões dos clientes, contribuindo também para a ampliação dos conhecimentos relacionados ao *marketing* digital, área de atuação da autora.

Diante da oportunidade de contribuição ao mercado brasileiro, o cumprimento do objetivo geral desta pesquisa auxilia a síntese das discussões disponíveis na literatura e sua aplicabilidade em *e-commerce* de varejo.

2 REVISÃO DA LITERATURA

Neste capítulo é apresentado o referencial teórico para fundamentar a pesquisa de acordo com os objetivos propostos. Sendo contemplado os temas: *e-commerce*, *marketing* digital, *customer churn*, mineração de dados, *machine learning* e análise preditiva.

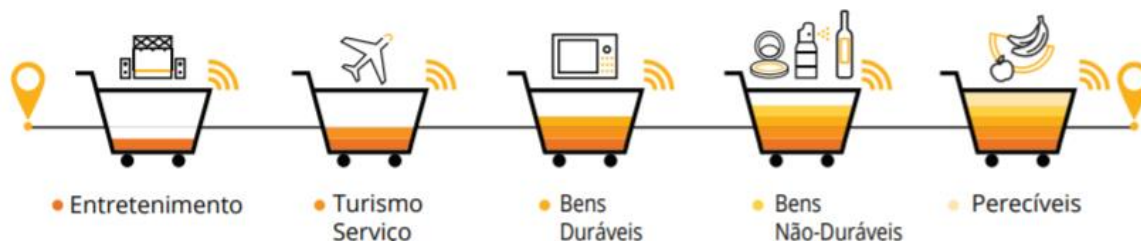
2.1 E-COMMERCE

O'Brien (2004, p.205) define *e-commerce* “como a compra e a venda por meios digitais.” Cardoso e Manganote (2000, p. 42) define como “a capacidade de realizar transações envolvendo a troca de bens ou serviços entre duas ou mais partes utilizando ferramentas eletrônicas e tecnologias emergentes”. Em 1979, Michael Aldrich inventou o primeiro sistema que permitia o processamento de transações online entre consumidores e empresa. Com a popularização da internet, em 1995 e o surgimento da Amazon.com, fundada por dois doutores em computação, as organizações passam a vender não apenas fisicamente, mas também virtualmente. Para Paraíso (2011) a principal vantagem dessa mudança para as empresas “é a eliminação de estoques, na medida em que repassa os pedidos de compra diretamente aos fornecedores, reduzindo substancialmente os custos em relação às lojas físicas”. No Brasil o setor começou a se desenvolver cinco anos depois, nos anos 2000, e desde então tem crescido. Nota-se que a comodidade, praticidade, inovação e conforto aos consumidores virtuais tem auxiliado o crescimento do *e-commerce*.

Segundo Ebit|Nielsen (2019), em um relatório sobre o comércio eletrônico brasileiro, após praticamente duas décadas e meia de atuação o mercado brasileiro em comércio eletrônico demonstra maturidade para um crescimento mais homogêneo e consistente, além de se encontrar no penúltimo nível de consumo, no que a empresa define como a rota global do *e-commerce*. A Ebit|Nielsen considera que a maior parte dos países tem uma rota de compra bem similar, que é a venda de entretenimento, depois turismo e serviços, bens duráveis, bens não duráveis e por último perecíveis (figura 3).

FIGURA 3: A ROTA GLOBAL DO E-COMMERCE

A ROTA GLOBAL DO E-COMMERCE



FONTE: EBIT | NIELSEN (2019)

Um dos maiores desafios para expansão do e-commerce brasileiro é a logística por conta de sua extensão, que afeta diretamente na variação da taxa de frete, além dos diferentes impostos, culturas, comunicação e comportamento de consumo em cada região. Mas ainda assim, em 2018 as vendas online registraram um faturamento de 12% a mais que em 2017, e a estimativa para 2019 é de aumento de 15% do faturamento (Ebit|Nielsen, 2019).

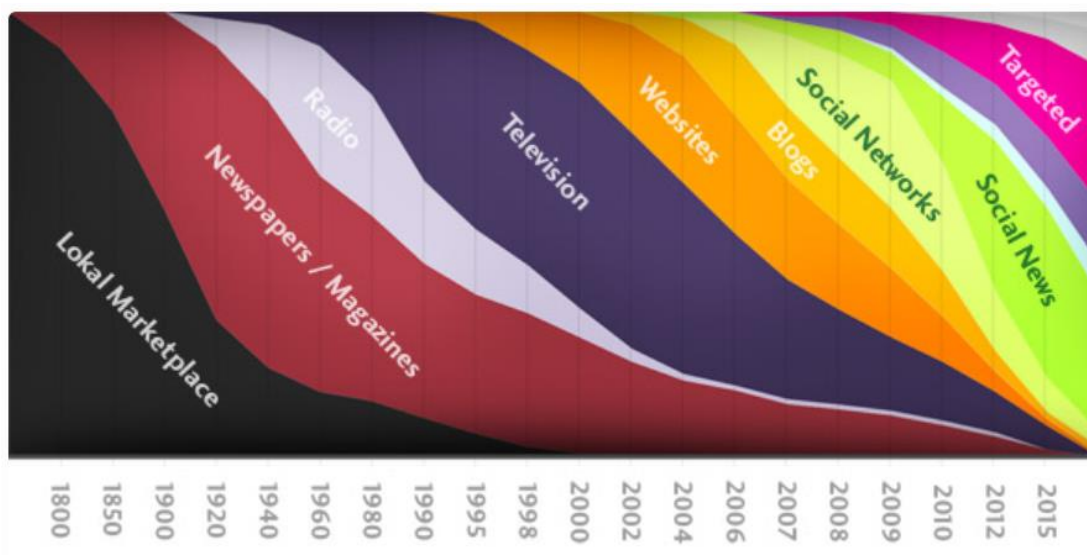
2.2 MARKETING DIGITAL

O *marketing* pode ser definido como um processo baseado na criação e oferta de produtos por alguém, de modo que gerem valor, para saciar as demandas de outras pessoas e companhias, por meio de uma troca (KOTLER, 2000). Ainda segundo Kotler há uma mudança no *marketing* com o passar dos anos, tendo passado do *marketing* centrado no produto, para voltado ao consumidor, depois para direcionado ao ser humano e com os avanços tecnológicos o comportamento do consumidor está sendo alterado novamente, e o *marketing* vem se tornando muito mais centrado no ser humano, dando muito mais poder ao consumidor conectado.

A transformação do comportamento do consumidor está relacionada com a fonte de informação dos indivíduos, Baekdal (2009) faz uma análise e apresenta algumas fases pelas quais houve transição na forma que as pessoas se comunicavam, conforme apresentado na figura 4. Segundo ele, em 1800 a comunicação era dada cara a cara, as interações aconteciam nos mercados locais,

e para ter acesso a informação as pessoas precisavam ir às ruas e conversar com outros indivíduos, o *marketing* era realizado boca a boca; em 1900 os jornais e revistas revolucionaram a forma de comunicação e busca de informação, já que a partir deles a informação ficava mais pulverizada, anúncios eram inseridos nas publicações impressas.

FIGURA 4: LINHA DO TEMPO DA INFORMAÇÃO



FONTE: BAEKDAL (2009)

A partir de 1920 a rádio começou a atrair a atenção das pessoas, e a partir de 1960 passou a dominar as fontes de informação com notícias ao vivo (BAEKDAL, 2009), o *marketing* era realizados por meio de anúncios em áudios, nesse período os *jingles*, que são músicas publicitárias surgem. Desde 1950 a televisão foi introduzida e em 1990 o público passou a ter grande interesse nas notícias assistidas (BAEKDAL, 2009). Os anúncios ganham recursos visuais, e algo muito utilizado eram as propagandas de respostas diretas (KOTLER, 2000), onde nos anúncios eram incluídos números de discagem gratuita para o consumidor efetuar a compra na hora. Em 1998 a *internet* se populariza, e as empresas começam a utilizá-la estrategicamente (BAEKDAL, 2009), logo, o *marketing* ganha um novo canal, com grandes vantagens como o custo baixo, não ser limitado por espaço de propaganda quando comparado às mídias impressas, televisão e o rádio, o acesso a recuperação das informações são rápidas e a compra pode ser feita com maior privacidade e rapidez (KOTLER, 2000, p.635).

Em 2004 as televisões e jornais ainda dominavam as fontes de notícias, mas a informação online já possuía grande relevância, qualquer pessoa com acesso a internet podia publicar e opinar, gerando uma sobrecarga de informação. As pessoas possuíam o suficiente, e não eram impactados com qualquer publicação e agora era necessário fazer algo melhor de forma diferente (BAEKDAL, 2009). A partir de 2007 diversos blogs surgiram e todos queriam informar e conectar seus pensamentos com outras pessoas, em 2009 surgem as redes sociais e todos estavam conectados e comunicando.

O *marketing* é composto por quatro elementos que são chamados de 4P's, Kotler (2015) define-os como "o conjunto de ferramentas que a empresa usa para atingir seus objetivos de *marketing* no mercado alvo". Sendo eles: Produto, Preço, Promoção e Praça ou Ponto de venda. Mais recentemente com o desenvolvimento das novas mídias, principalmente as digitais, surgiram conceitos novos que procuram adaptar a idéia dentro da realidade do *marketing* digital, conforme afirma Turchi (2012, p. 1) "o mundo digital tem trazido mudanças profundas ao mercado e obrigado todos os profissionais a se reajustarem rapidamente diante desta nova realidade". Por tanto utilizam ferramentas como as redes sociais (Facebook, Instagram, Twitter); as tecnologias de SEO dentro de buscadores (Google, Bing, etc.), blogs, e mails, realidade aumentada e QR code. Dentro desta nova realidade Vaz (2011) traz um método que quando aplicado permite que a empresa aprenda sobre seu mercado e por ser um processo permite que seja melhorado seu desempenho a cada ação, esta proposta foi denominada os 8Ps do *marketing* digital, sendo eles:

- pesquisa: consiste na busca de informações sobre o consumidor e seus hábitos na *internet*, ela é baseada na experiência de consumo dos clientes e na forma como interagem;
- planejamento: deve-se elaborar as ações e os passos a serem traçados ao longo de todo o ciclo estratégico, sendo necessária a produção de um documento contendo todo o planejamento do *marketing* digital;
- produção: responsável pela produção e entrega efetiva do site;
- publicação: é todo conteúdo gerado para relacionamento com o cliente;
- promoção: é o modo como divulga-se a marca para o público alvo desejado tendo como objetivo o alcance do resultado final que pode ser a venda de produtos;

- propagação: é a propagação da marca para o maior número de pessoas possível, sendo importante a comunicação viral.
- personalização: consiste no tratamento do cliente de forma personalizada, para Vaz (2011) quanto mais segmentada a base de clientes mais fácil tratar de forma personalizada cada cliente;
- precisão: mensurar os resultados obtidos nos “Ps” anteriores e com esta informação reiniciar o processo.

2.3 CUSTOMER *CHURN*

O termo *churn* é utilizado na indústria e na literatura para referir-se à descontinuação de um contrato (Lazarov & Capota, 2007). O *churn* pode ser medido através do percentual de clientes que uma empresa perde num período de tempo específico, geralmente um ano (Lejeune, 2001). Gladly et al. (2009), usando uma abordagem centrada no cliente, definiram *churn* como aquele cliente cujo ciclo de vida é decrescente ao longo do tempo.

Com o saturamento do mercado, os consumidores adquirem maior poder de escolha e mudam suas compras com base em melhores preços e serviços, considerando esse fato as organizações têm dado maior atenção aos clientes, e tentam mantê-los e fazê-los efetivar a compra.

Há dois tipos de *churn*: involuntário e o voluntário (Hadden et al., 2007). O involuntário é quando ele não paga e a empresa deixa de oferecer os produtos ou serviços; e o voluntário é quando o cliente decide não efetuar a compra ou cortar o relacionamento com a empresa, este se divide em acidental, em que o cliente Não consegue por algum motivo manter o relacionamento com a empresa, e o deliberado que é quando o cliente toma a decisão de trocar de fornecedor. A retenção de clientes tem um grande valor pois é mais rentável manter os atuais clientes do que adquirir novos. Para Neslin et al, (2006), uma forma de gerenciar o *churn* é prever quais clientes com mais probabilidade de abandonar o relacionamento com a empresa e trabalhar com eles a fim de tentar evitar a ocorrência desse rompimento. Para calcular a taxa de *churn*, é a percentagem de clientes que não fazem mais compras dentro de um determinado prazo.

2.4 MINERAÇÃO DE DADOS

Antes de entender o conceito de mineração de dados é necessário comentar sobre o processo de descoberta de conhecimento (*Knowledge Discovery in Databases* - KDD). O KDD tem como enfoque a descoberta de conhecimento a partir de dados. O KDD surgiu com o objetivo de ajudar os humanos em sua extração de informações úteis dos grandes volumes de dados (*Fayyad et al*, 1996). O processo geral do KDD (figura 5) inclui a avaliação e possível interpretação dos padrões minerados para determinar quais podem ser considerados novo conhecimento.

FIGURA 5: ETAPAS DO PROCESSO DE *KNOWLEDGE DISCOVERY IN DATABASES* (KDD)



FONTE: ADAPTADO DE FAYYAD ET AL. (1996, P. 41).

Resumidamente as etapas podem ser descritas como:

- seleção: escolha de um conjunto de dados ou concentrando-se em um subconjunto de variáveis ou amostras de dados, nas quais pretende-se descobrir conhecimentos;
- limpeza e pré-processamento: remoção de ruído, coleta das informações necessárias para modelá-lo ou explicá-lo, decisão de estratégias para lidar com campos de dados ausentes;
- transformação: representação dos dados com recursos úteis de acordo com o objetivo da tarefa, utilizando métodos de redução ou transformação;
- mineração de dados: combinação dos objetivos do processo de KDD para minerar os dados utilizando métodos e algoritmos específicos, procurando

por padrões de interesse, esse tópico será melhor definido posteriormente nesta sessão;

- interpretação e avaliação: visualização e interpretação dos padrões obtidos gerando um novo conhecimento descoberto.

O KDD refere-se a todo o processo de descobrimento do conhecimento, enquanto *data mining* ou mineração de dados refere-se a um passo do processo que consiste na aplicação de análise sobre os dados a partir de algoritmos que respeitem as limitações da capacidade computacional produzindo assim padrões (CAMILO E SILVA, 2009).

A mineração de dados é uma disciplina interdisciplinar e multidisciplinar que envolve conhecimento de áreas como banco de dados, estatística, aprendizagem de máquina, computação de alto desempenho, reconhecimento de padrões, computação natural, visualização de dados, recuperação de informação, processamento de imagens e de sinais e inteligência artificial. Pode ter tarefas categorizadas em descritivas e preditivas, a primeira caracteriza as propriedades dos dados e a segunda faz sua conclusão a partir dos dados com o objetivo de predição (CASTRO E FERRARI, 2016). No Data Mining são utilizadas ferramentas que podem substituir e/ou aprimorar a inteligência humana, pois as ferramentas são capazes de analisar grande volume de dados (Pereira, 2015, p.26).

Análise descritiva dos dados utiliza ferramentas capazes de medir, explorar e descrever características dos dados, permitindo investigar a distribuição de frequência, as medidas de centro e variação, e as medidas de posição relativa e associação dos dados. Já a análise preditiva permite verificar a consequência das informações geradas a partir da base de dados minerada, falaremos mais sobre esse tema na seção 2.7.

Boente, Goldschmidt e Estrela (2008) descrevem que existem diferentes tarefas de Mineração de Dados para suprir as diferentes necessidades relacionadas à análise de diferentes tipos de dados:

- regras de associação: visa detectar correlações entre diferentes itens em um conjunto de dados;
- classificação: é a tarefa de atribuir determinados objetos a uma categoria predefinida;
- regressão: consiste na busca por funções que possam descrever determinadas relações entre variáveis de um banco de dados;

- clusterização: consiste na separação de objetos em grupos (clusters) semelhantes. (BOENTE; GOLDSCHMIDT; ESTRELA, 2008).

2.5 MACHINE LEARNING

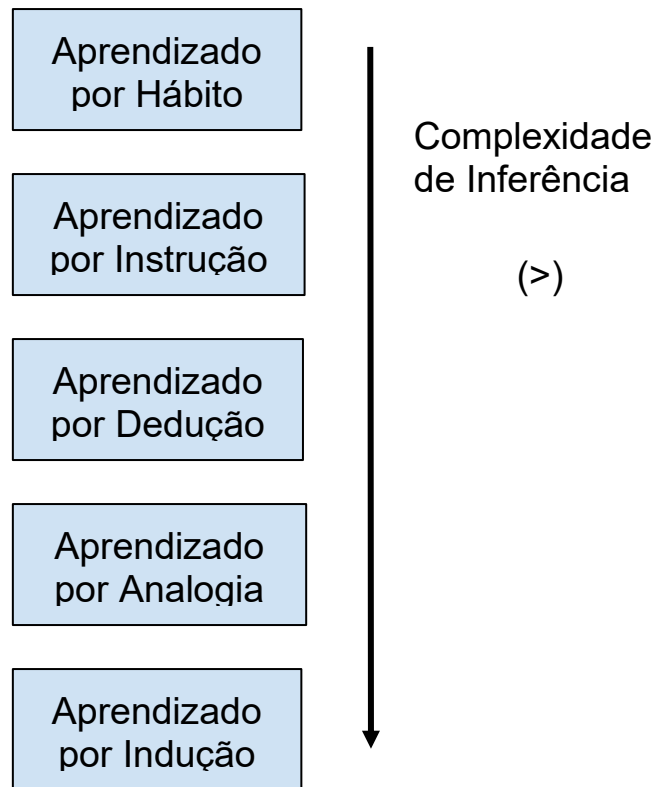
Aprendizado de Máquina (*machine learning*) é uma área da inteligência artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado e a construção de sistemas capazes de adquirir conhecimentos de forma automática (MONARD, BARANAUSKAS, 2003). De acordo com Mitchell (1997), algoritmos de *machine learning* têm sido de grande valor prático para diversas aplicações, como, problemas de Mineração de Dados (Data Mining), onde grandes banco de dados são analisados automaticamente, na busca de padrões que possam ser úteis; onde humanos não poderiam ter o conhecimento necessário para desenvolvimento de algoritmos efetivos; o custo da aquisição ou codificação manual do conhecimento é grande; e programa necessita se adaptar dinamicamente a mudanças.

Uma perspectiva útil na aprendizagem de máquina é que ela envolve a busca de um grande espaço de hipóteses possíveis para determinar aquela que melhor se adapta aos dados observados e qualquer conhecimento prévio.

Ainda que o *machine learning* seja uma ótima ferramenta para aquisição de conhecimento, não existe algoritmo que apresente o melhor desempenho para todos os problemas, deste modo, é necessário entender as limitações de cada um deles utilizando metodologias para avaliação.

Existem algumas estratégias de aprendizado básicas relatadas para qualquer modo de aquisição do conhecimento, sendo elas: por hábito, instrução, dedução, analogia e indução (Figura 6).

FIGURA 6: CLASSIFICAÇÃO DAS ESTRATÉGIAS DE APRENDIZADO DE MÁQUINA



FONTE: MONARD ET. AL (2003)

No aprendizado por hábito o conhecimento é adquirido diretamente pelo aprendiz sem nenhuma inferência, esta estratégia é utilizada para reconhecer um conceito por meio de um algoritmo específico, ou construído uma base de dados fatuais sobre o conceito; no aprendizado por instrução o aprendiz assimila conceitos de uma fonte não direta, as informações são analisadas e são selecionadas os fatos mais relevantes; Já no aprendizado por dedução o conhecimento é adquirido por meio de dedução sobre o conceito já adquirido; No aprendizado por analogia um novo conhecimento é obtido ao modificar a definição de um conceito semelhante já conhecido, podendo ser visto como uma combinação entre o aprendizado indutivo e dedutivo, e; O aprendizado por indução se trata do conhecimento adquirido por inferências indutivas sobre o fato apresentado, onde hipóteses são geradas, podendo ou não ser preservadas a veracidade.

Sendo assim, o aprendizado possui uma hierarquia, a indução é a forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto, partindo da parte para o todo, permitindo a geração de um novo conhecimento e predição de eventos futuros, mesmo tendo um risco de não preservar a verdade.

Pode realizar-se de forma supervisionada (*supervised*) ou não supervisionada (*unsupervised*). *Supervised machine learning* refere-se à procura de algoritmos que aprendam através de instâncias externas de forma a produzirem hipóteses, que posteriormente fazem previsões sobre instâncias futuras (Kotsiantis, 2007), onde cada exemplo é descrito por um vetor de atributos. O objetivo do algoritmo de indução é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados. *Unsupervised learning* necessita descobrir as relações, padrões, regularidades ou categorias nos dados que lhe vão sendo apresentados e codificá-las nas saídas (Rojas, 1996), o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou clusters (Cheeseman & Stutz 1990). Depois é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema.

Para avaliar o desempenho dos métodos, e comparar os algoritmos a principal metodologia de avaliação utilizada é a de amostragem (*resampling*), onde há um conjunto de dados de exemplo, que tenha um tamanho finito e um indutor, o objetivo de um indutor consiste em extrair um bom classificador a partir de um conjunto de exemplos rotulados. É importante que a amostra seja aleatória, ou seja, que não haja pré-seleção. Para problemas reais, normalmente é tomada uma amostra de tamanho n e o objetivo consiste em estimar uma medida para aquela população em particular.

2.6 ANÁLISE PREDITIVA

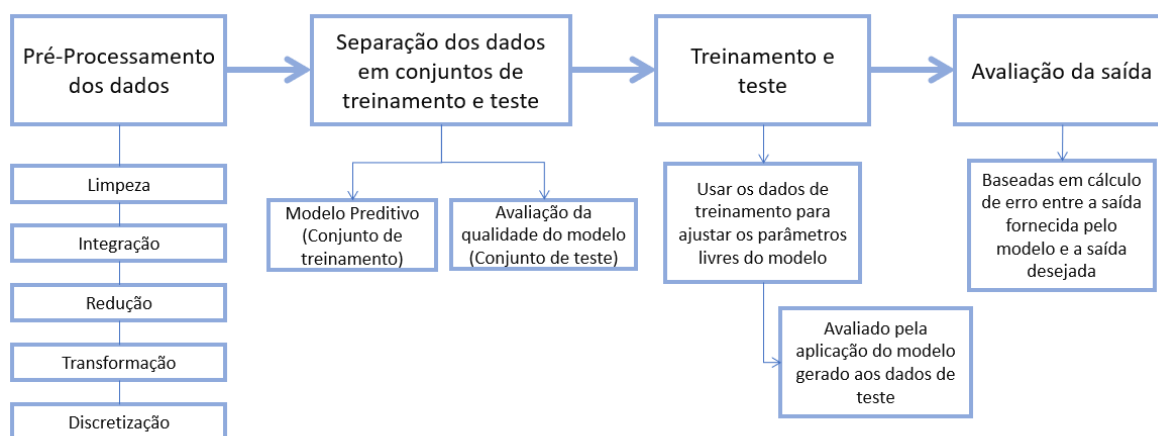
Análise preditiva é o ramo da mineração de dados que ajuda a prever as tendências e a estimar as probabilidades de que eventos ocorrerem. Para cada resultado de interação do cliente com a compra de um produto em um *e-commerce*, há um conjunto de ações que possuem um registro histórico. Com isso é possível agrupar os dados em grupos similares para produzir uma taxonomia ou classificação. O agrupamento de dados, é um termo usado para a utilização de métodos numéricos de análise de dados com o objetivo de descobrir grupos homogêneos de objetos. A demanda para predição de acontecimentos surgiu com uma falta que os sistemas *business intelligence* (BI) tinham, já que ajudavam os

executivos a entender o que aconteceu, mas havia a necessidade de ferramentas que predissesse o que iria acontecer e para onde o seu negócio estava indo.

Sempre que houver um conjunto de objetos que possa ser rotulado e usado para treinar um modelo de forma supervisionada é possível projetar um modelo preditivo capaz de estimar o valor de novos objetos (Castro e Ferrari, 2016). Esse modelo preditivo pode ser dividido em quatro etapas (figura 7):

- pré-processamento dos dados: consiste na preparação da base de dados;
- separação dos dados: os dados disponíveis podem ser separados em duas partes, uma para uso no conjunto de treinamentos, ou seja, é usado na geração do modelo preditivo, e a outra parte no conjunto de teste, onde se avalia a qualidade do modelo gerado;
- treinamento e teste: o treinamento consiste em usar os dados de treinamento para ajustar os parâmetros livres do modelo;
- avaliação da saída: por se tratar de um treinamento do tipo supervisionado a avaliação da saída é baseada em cálculos de erro entre saída fornecida e a saída desejada.

FIGURA 7: RESUMO DAS ETAPAS DO PROCESSO DE UM MODELO PREDITIVO



FONTE: ADAPTADO DE CASTRO E FERREIRA (2016, P. 256)

A predição é dividida em duas tarefas: classificação e estimação. Na classificação cada objeto da base possui a correspondente classe à qual pertence, sendo assim, a tarefa do algoritmo de classificação é identificar a classe à qual pertence um novo objeto ainda não apresentado. Enquanto a estimação corresponde à tarefa de predição de um valor contínuo de uma variável, onde se

requer pares entrada-saída desejada para a construção do estimador e possui muitas características e processos em comum com a classificação.

2.7 RELAÇÃO ENTRE MACHINE LEARNING E *CHURN* PREDICTION

Mofakham et al (2008) modela promoções para novos clientes em um *e-commerce*. Primeiro introduziram o modelo de agrupamento com capacidade de aprendizagem para formular promoções para novos clientes. Neste algoritmo de aprendizado modificado, no lugar de uma categoria de um único produto, o vendedor envia ao novo cliente uma promoção em um conjunto variável de categorias com a maior classificação de desejo entre os clientes anteriores similares, diminuindo assim a chance de *churn*.

Yu et al (2011) propõem máquina de vetores de suporte estendido (ESVM) pela introdução de parâmetros para informar o impacto de *churner*, *non-churner* e *nonlinear*. Utilizam os algoritmos de predição rede neural artificial (RNA), árvore de decisão, SVM e ESVM. O resultado mostra que o ESVM apresenta o melhor desempenho entre eles nos aspectos de precisão, taxa de acerto, taxa de cobertura, coeficiente de sustentação e tempo de tratamento.

Shim et al (2012) propõe estratégias de CRM para *e-commerces* de pequeno porte baseados em regras de associações e padrões sequenciais, em primeiro momento classificou os clientes VIPs de acordo com recência, frequência e monetária (RFM), então, desenvolveram um modelo que classifica os clientes em VIP ou Não -VIP, usando várias técnicas de mineração de dados (árvore de decisão, rede neural artificial, regressão logística e ensacamento).

Kim e Lee (2012) propõe um método que se inicia particionando um conjunto de dados em subconjuntos e aplica o aprendizado múltiplo sequencial para reduzir recursos e fornecer resultados consistentes para subconjuntos de dados combinados. CHEN, Zhen-Yu; FAN, Zhi-Ping; SUN, Minghe, (2012) utiliza vetores de suporte a vários núcleos hierárquicos (H-MK-SVM), onde desenvolveu um algoritmo de treinamento trifásico. O H-MK-SVM constrói uma função de classificação estimando os coeficientes de variáveis comportamentais estáticas e longitudinais no processo de treinamento sem a transformação dos dados longitudinais comportamentais.

Chen (2012) propõe um novo mecanismo baseado em gráfico gama CUSUM em que apenas o tempo de chegada (IAT) e a recência precisam ser coletados, de modo que os parâmetros personalizados podem ser estimados para fins de monitoramento individual. Keramati et al (2016) utiliza árvore de decisão para previsão de *churn* em um banco eletrônico. Wu e Meng (2016) utiliza SMOTE e no *AdaBoost* aprimorados para previsão de *churn* em um *e-commerce* B2C. Primeiro, processaram os dados de rotatividade com SMOTE aprimorado, que combina métodos de *oversampling* e *undersampling* para resolver o problema de desequilíbrio e, posteriormente, integraram o algoritmo *AdaBoost* para prever o *churn*.

Chamberlain et al (2017) descrevem um sistema de previsão de *Customer Life Time Value* (CLTV) implantado em uma varejista de moda online onde utiliza rede neural antecipada nas características de uma sessão supervisionada; aprendendo uma incorporação de clientes usando dados da sessão em um sem supervisão para aumentar o conjunto de recursos. Subramanya e Somani (2017) utiliza regressão logística regularizada, SVM e Gradient Boost *Random Forests* para *predição de churn*.

Valero Fernandes et. Al (2017) utiliza uma variedade de algoritmos para classificações baseadas em comportamentos específicos e focados no marketing do cliente de um *e-commerce* do Reino Unido. Os modelos implementados (e comparados) para classificação foram: Regressão Logística, Análise Discriminante Quadrática, SVM Linear, SVM RBF, Processo Gaussiano, Árvore de Decisão, *Random Forest* e *Perceptron Multicamadas* (Rede Neural).

Rachid et al (2018) propuseram um modelo de predição de *churn* em *e-commerce*, onde a clusterização foi baseada na integração do método *k-means* e do modelo *Length-Recency-Frequency-Monetary* (LRFM), após, fizeram a predição multiclasse baseada em três técnicas de classificação: árvore de decisão simples, redes neurais artificiais e conjunto de árvores de decisão, em que se classifica um cliente de acordo com padrões de compra como: fiéis (*non-churned*), deflector parcial (*partially-churned*) e desertor total (*totally-churned*). No mesmo ano Agrawal et al (2018) utiliza uma abordagem de *deep learning* para a previsão de rotatividade em um conjunto de dados da Telco, primeiro utiliza rede neural multicamadas para classificação. O modelo de previsão de rotatividade funciona nos recursos do cliente, recursos de suporte, recursos de uso e recursos contextuais, já o modelo

treinado aplica os pesos finais sobre esses recursos e prevê a possibilidade de rotatividade para esse cliente.

Wang, Hussain e Xu (2018) utilizam árvore de decisão de aumento de gradiente (GBDT) para prever se um cliente será um *churn* baseado em suas atividades nos anúncios de busca. Foram extraídos dois tipos de recursos para o GBDT: recursos dinâmicos, ou seja, sequência de atividades dos clientes (por exemplo, impressões, cliques) durante um longo período, e recursos estáticos, onde consideraram as informações de configuração de clientes (por exemplo, tempo de criação, tipo de cliente).

3 METODOLOGIA

Silva e Menezes (2005) caracteriza pesquisa como “um conjunto de ações, propostas para encontrar a solução para um problema, que têm por base procedimentos racionais e sistemáticos”. Após a definição do problema de pesquisa é necessário delimitar qual será o caminho percorrido pelo pesquisador para garantir o atingimento dos objetivos traçados.

3.1 CARACTERIZAÇÃO DA PESQUISA

Esta pesquisa é caracterizada como quantitativa, onde pode-se transformar opiniões e informação em números para então poder classificá-las e analisá-las através de técnicas estatísticas (SILVA e MENEZES, 2005). Quanto aos objetivos proposta por Gil (2007): pesquisa exploratória, pesquisa descritiva e pesquisa explicativa. Com base nessa classificação, esta pesquisa define-se como exploratória. Para Malhotra (2006), o objetivo principal da pesquisa exploratória é possibilitar a compreensão do problema enfrentado pelo pesquisador. Tendo como propósito de proporcionar maior familiaridade com o tema, podendo envolver levantamento bibliográfico, entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado, e análise de exemplos que estimulem a compreensão (GIL, 2007).

Quanto aos procedimentos que se referem a maneira que se conduz o estudo e se obtém os dados. Se destacando os tipos: estudo de caso, pesquisa de levantamento, pesquisa bibliográfica, pesquisa documental, pesquisa experimental e pesquisa participante (GIL, 2007). Esta pesquisa trata-se de uma pesquisa bibliográfica Fonseca (2002, p.32) a caracteriza como:

A pesquisa bibliográfica é feita a partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de web sites. Qualquer trabalho científico inicia-se com uma pesquisa bibliográfica, que permite ao pesquisador conhecer o que já se estudou sobre o assunto. Existem, porém, pesquisas científicas que se baseiam unicamente na pesquisa bibliográfica, procurando referências teóricas publicadas com o objetivo de recolher informações ou conhecimentos prévios sobre o problema a respeito do qual se procura a resposta. (FONSECA, 2002, p.32)

Além de ser caracterizada como uma Pesquisa Experimental, que consiste em um procedimento metodológico onde as variáveis são controláveis. Assim, a lógica desta pesquisa compreende uma ação ativa pesquisador na análise dos efeitos das variáveis sobre o objeto.

Sendo assim, essa pesquisa irá analisar as publicações existentes sobre análise preditiva de *churn* e reunir conhecimento voltado ao uso de *machine learning* dentro da área do *marketing* digital e aplicará os métodos levantados em uma base de dados real, para analisar os resultados. A figura 8 apresenta a síntese da caracterização dessa pesquisa.

FIGURA 8: CARACTERIZAÇÃO DA PESQUISA

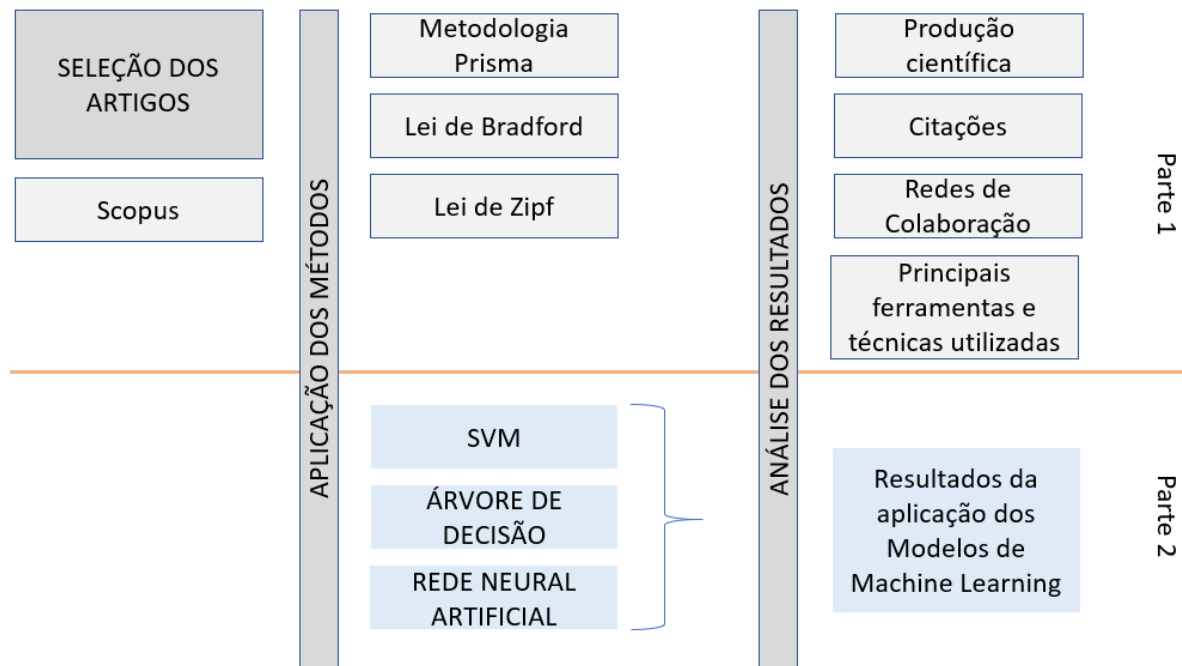
Abordagem	Objetivos	Procedimentos
<input type="checkbox"/> Qualitativa	<input type="checkbox"/> Descritiva	<input type="checkbox"/> Estudo de caso
<input checked="" type="checkbox"/> Quantitativa	<input type="checkbox"/> Explicativa	<input type="checkbox"/> Pesquisa de levantamento
	<input checked="" type="checkbox"/> Exploratória	<input checked="" type="checkbox"/> Pesquisa bibliográfica
		<input type="checkbox"/> Pesquisa documental
		<input checked="" type="checkbox"/> Pesquisa Experimental
		<input type="checkbox"/> Pesquisa Participante

FONTE: A AUTORA (2019).

3.2 ETAPAS, MATERIAIS E MÉTODOS

A pesquisa será dividida em duas partes, sendo a primeira com três etapas principais, a primeira será a seleção dos artigos nas bases de dados, a segunda a aplicação dos métodos de análise bibliométricos, infométricos e cientométricos e a terceira etapa será a análise dos resultados. Após a análise bibliográfica, inicia-se a segunda etapa que consiste na aplicação dos principais métodos de *machine learning* levantados na parte 2, e posteriormente a análise dos resultados (figura 9).

FIGURA 9: ETAPAS DA REALIZAÇÃO DA PESQUISA



FONTE: A AUTORA (2019).

Os métodos que serão utilizados incluem a aplicação técnicas matemáticas e estatísticas para descrever aspectos da literatura, que incluem a aplicação das fórmulas referentes às Lei de Bradford.

A lei de Bradford possui foco na dispersão de trabalhos sobre um determinado assunto em periódicos, visa atenção à produtividade de periódicos, zonas de densidade variável quanto à concentração de trabalhos num conjunto de fontes, estimando o grau de relevância de periódicos sobre um determinado tema. Propõe que se dispormos periódicos em ordem decrescente de produtividade de artigos sobre um determinado tema, pode-se distinguir um núcleo de periódicos mais particularmente devotados ao tema e vários grupos ou zonas que incluem o mesmo número de artigos que o núcleo, sempre que o número de periódicos existentes no núcleo e nas zonas sucessivas seja de ordem de 1: n: n2: n3. representado por:

$$mb = [(z2/z1) + (z3/z2)]/2$$

Onde:

- mb = multiplicador de Bradford
- z1 = quantidade de periódicos da zona 1
- z2 = quantidade de periódicos da zona 2

- z_3 = quantidade de periódicos da zona 3

A lei de zipf possui foco na distribuição de palavras num texto. Preocupa-se com a frequência de palavras e a caracterização de regularidades associadas à linguagem, representada por:

$$l_1/l_n = n(n+1)/c$$

Onde:

- l_1 = número de palavras com ocorrência 1
- l_n = número de palavras com ocorrência n
- c = uma constante de acordo com o idioma

Após a análise bibliográfica foi realizado uma segunda fase da pesquisa, selecionados os modelos de *machine learning* mais utilizados foi aplicado e avaliado os resultados de cada um, utilizando uma base real.

Dos materiais necessários incluem as bases indexadoras de artigos e os software para processamento dos dados coletados. Para a escolha das bases de dados que serão utilizados alguns critérios, sendo eles:

- a disponibilidade de acesso das bases indexadoras pela Universidade Federal do Paraná;
- a acessibilidade a periódicos nacionais e principalmente aos internacionais por se tratar de um conteúdo recente e com pouca publicação nacional, como levantado na justificativa científica da pesquisa;
- os conteúdos devem possuir confiabilidade, possuindo a fonte de informação e conteúdo possui credibilidade;
- possuir publicações recentes e frequência de publicação.

Foram selecionadas a base de informação Scopus, que consiste em uma base de dados internacional privada pertencente ao grupo Elsevier, que disponibiliza acesso a jornais, livros, conferências e artigos de periódicos de diversas áreas do conhecimento.

Os softwares a serem utilizados serão:

- bibliometrix um pacote R para análise abrangente de mapeamento científico de dados bibliográficos das bases de dados SCOPUS, Web of Science, PubMed e Cochrane do Clarivate Analytics, realizando análises bibliométricas e construindo matrizes de dados para co-citação, acoplamento, análise de colaboração científica e análise conjunta;
- Weka 3.8, que é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Ele contém ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização.

3.3 METODOLOGIA PRISMA

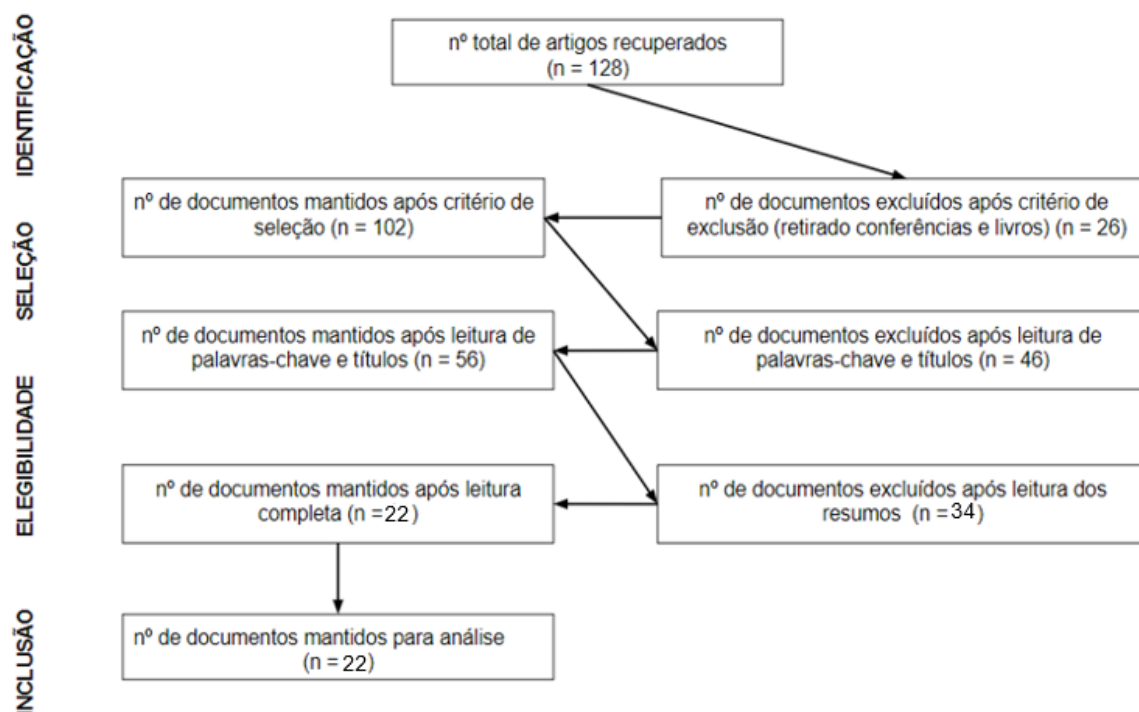
O PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) é uma diretriz proposta por Moher et al. (2009) que possui o objetivo de auxiliar a melhoria da qualidade do relato dos dados da meta-análise. Surgiu de uma revisão e atualização do QUOROM – (Quality Of Reporting Of Meta-analyses), a principal razão da mudança de nome foi devido à necessidade de se diferenciar RS de metanálise. O PRISMA é composto de um checklist de 27 itens e um diagrama de fluxo de seleção de artigos de quatro fases (Identificação, Seleção, Elegibilidade e Inclusão) que foram adaptadas para esta pesquisa.

Para a primeira etapa, a de Identificação, foi realizada uma busca na base de dados Scopus, utilizando os termos "*machine learning*" AND "*E-commerce*" AND "*Churn prediction*", excluíram-se livros, revistas e publicações comerciais por em sua maioria não disponibilizar o acesso ao texto completo, consideram-se ainda todos os idiomas e não houve recorte de período.

Foram retornados 128 documentos, que após a exclusão de conferências, livros e revisões, restaram 102 para a etapa de seleção/elegibilidade. Após a leitura dos títulos, palavras-chave e resumos, foram mantidos 56 documentos, os documentos excluídos nesta etapa apontavam outras áreas de atuação, em sua maioria telecomunicações. Na etapa de elegibilidade restaram 56 para a leitura dos resumos dos quais foram mantidos 22 para a leitura completa. As exclusões destes 34 artigos foram devidas a estarem relacionados a outra área como jogos digitais, transporte, B2B, Segurança, Mineração de Texto, Financeiro, Gestão do

conhecimento, Viagens e Jogos Online. Foi incluído 22 documentos para análise dos resultados. A Figura 10 ilustra as etapas adotadas na criação da base de documentos analisados no estudo.

FIGURA 10: METODOLOGIA PRISMA ADAPTADA



FONTE: A AUTORA (2019).

3.4 BASE DE DADOS E APLICAÇÃO

Foi utilizado uma base de dados de um e-commerce real, na área de varejo, um dos maiores marketplaces de produtos para casa do Brasil. Fundado em 2009, no início vendendo apenas pisos de madeira e revestimentos, hoje conta com diversas categorias e mais de um milhão de produtos. Em 2016, o alto faturamento os colocou como 21º maior comércio eletrônico do país, de acordo com o ranking da Sociedade Brasileira de Varejo e Consumo (SBVC), e em 2018 inaugurou o *marketplace*, onde diversos *sellers* podem vender em sua loja virtual. Conta com investidores que incentivam o desenvolvimento tecnológico. A empresa não se intitula apenas uma varejista, mas uma empresa de tecnologia. Em seus 10 anos de história desenvolveu mais de 20 softwares focando na experiência do cliente. A organização trabalha com o modelo de negócio chamado *dropshipping*, pois, não trabalham com estoques, e sim com fornecedores de quem eles compram os

produtos e transportadoras parceiras que pegam o item direto no fornecedor e levam ao consumidor.

Há uma segmentação de clientes conforme comportamento de compra no decorrer de um ano móvel que é atualizado utilizando K-Means, um método de clusterização que objetiva particionar n observações dentre k grupos onde cada observação pertence ao grupo mais próximo da média, dividindo os clientes em três grupos: Ouro, Prata e Bronze. Sendo que Ouro define os clientes que possuem uma frequência de compra maior e ticket médio mais alto, Prata os compradores frequentes com ticket médios menores e os Bronze os clientes que compram numa frequência menor, porém possuem mais de uma compra.

Para aplicação dos métodos selecionados após a análise bibliográfica, e apontado os três principais para teste na base, será utilizado o Weka 3.8, que é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Ele contém ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização. A tabela 1 apresenta a relação dos atributos, tipo e descrição de cada um deles.

TABELA 2 - DESCRIÇÃO DA BASE DE DADOS

Nome do Atributo	Tipo	Descrição
idcli	Numérica	ID de cada cliente
tipopessoa	Nominal	Pessoa Física ou Jurídica
estado	Nominal	Estado em que o cliente mora
newsletter	Binário	Cliente aceitou receber emails com promoções
valorgasto	Numérica	Soma do valor absoluto comprado pelo cliente
pedidos	Numérica	Quantidade de pedidos realizados pelo cliente
taxaaprovação	Numérica	Taxa de aprovação do cliente (valor realmente pago/valor comprado)
cadastroprimeiracompra	Numérica	Diferença de dias entre o cadastro e a primeira compra
primeiraultimacompra	Numérica	Diferença de dias entre a primeira compra e a última compra
churn	Nominal	Definição do cliente entre chun (Sim) ou não Churn (Não)

FONTE: A AUTORA (2019)

3.5 MÉTODOS DE MACHINE LEARNING

Após análise bibliográfica, detalhada no próximo capítulo, foram selecionados três métodos de *machine learning* para teste: C4.5, RNA e SVM, que serão explicados neste capítulo.

3.5.1 ÁRVORE DE DECISÃO (J48)

As árvores de decisão consistem de nodos que representam os atributos, e de folhas, que representam as diferentes classes de um conjunto de treinamento (Holsheimer e Siebes, 1994). São utilizadas em algoritmos de classificação, auxiliando na construção de classificadores que predizem classes baseadas nos valores dos atributos.

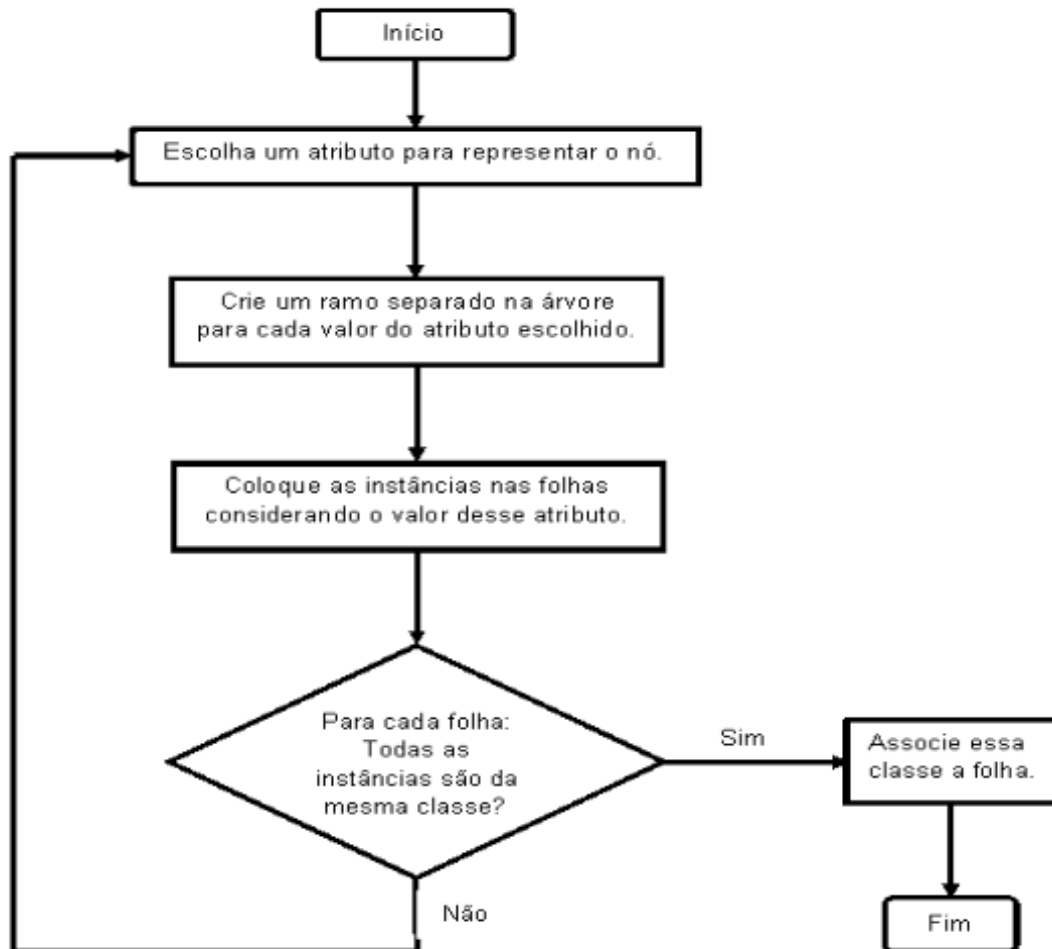
Para classificação de um objeto, se inicia pela raiz da árvore, aplica-se o teste em cada nodo e retorna-se o ramo apropriado para a saída. E esse processo continua até que seja encontrada o objeto classificado segundo a classe indicada naquela folha.

Uma árvore de decisão participa de um conjunto de treinamento, até que cada subconjunto obtido tenha casos de uma única classe, para isso, é examinado e comparado a distribuição de classes durante a construção da árvore.

Os algoritmos mais utilizados de árvores de decisão são o ID3 e o C4.5, que foram introduzidos por Quinlan para indução de modelos de classificação. Um dos primeiros algoritmos, foi o ID3, foi elaborada baseada em sistemas de inferência e em conceitos de sistemas de aprendizagem, construindo árvores a partir de um conjunto de exemplos. Ele separa um conjunto de treinamento em subconjuntos, até que contenham exemplos de uma única classe, sendo a divisão realizada a partir de um único atributo, que é selecionado por meio de uma propriedade estatística, denominada ganho de informação, que mede quanto informativo é um atributo.

O algoritmo C4.5 é uma evolução do algoritmo ID3, pois trabalha com valores indisponíveis, contínuos, poda de árvores de decisão e derivação regras. Existe um mapeamento entre cada objeto associado aos resultados, onde cada subconjunto pode ser substituído por uma árvore de decisão, o resultado será uma árvore para todos os elementos. Deste modo, um teste que gera uma divisão não trivial de qualquer conjunto de objetos sempre pode ser encontrado, este procedimento produzir uma árvore de decisão que classifica corretamente os objetos. O algoritmo C4.5 examina todos os atributos candidatos e escolhe a que maximiza o ganho de informação. O processo é repetido recursivamente para obter os demais nós e formar a árvore de decisão com os subconjuntos restante, a figura 11 apresenta o fluxograma do algoritmo C4.8.

FIGURA 11: FLUXOGRAMA DO ALGORITMO DE ÁRVORE DE DECISÃO



FONTE: QUINLAN, 1993 APUD RUIZ ET AL, 2009

3.5.2 REDE NEURAL ARTIFICIAL (RNA)

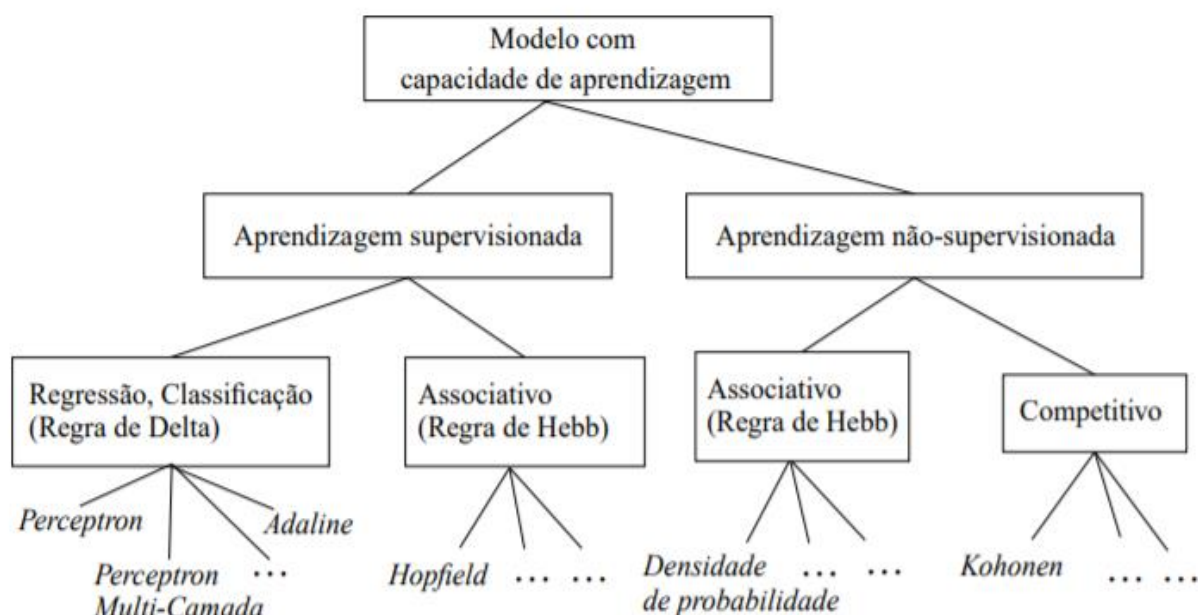
As redes neurais artificiais (RNAs) são sistemas massivos e paralelos distribuídos, compostos de unidades simples de processamento, que calculam funções matemáticas, sendo ligadas por um grande número de conexões. O funcionamento dessas redes é inspirado no cérebro humano (Braga et al., 2000). Possuem a capacidade de extrair características não explícitas de um conjunto de informações que são fornecidas (Kovács, 1996).

A maior vantagem das RNAs sobre os métodos convencionais é que elas não requerem informação detalhada sobre os processos físicos do sistema a ser modelado, sendo este descrito no modelo de entrada-saída. Uma das propriedades

mais importantes de uma RNA é a capacidade de aprender por meio de exemplos e fazer inferências sobre o que aprendeu, melhorando gradativamente o seu desempenho. As redes neurais utilizam um algoritmo de aprendizagem cuja tarefa é ajustar os pesos de suas conexões (BRAGA, et al, 2000).

Uma RNA tem duas facetas elementares, a arquitetura e o algoritmo de aprendizagem. Essa divisão surge naturalmente pelo paradigma como a rede é treinada. A rede é treinada por exemplos de treino. Existem duas formas básicas de aprendizado de RNAs: o aprendizado supervisionado onde um agente externo apresenta à rede neural alguns conjuntos de padrões de entrada e seus correspondentes padrões de saída, e; o não-supervisionado onde não existe um agente externo para acompanhar o processo de aprendizado. A rede processa as entradas e, detectando suas regularidades, tenta progressivamente estabelecer representações internas para codificar características e classificá-las automaticamente. A figura 12 apresenta a classificação estrutural e funcional de RNAs.

FIGURA 12: CLASSIFICAÇÃO ESTRUTURAL E FUNCIONAL DE REDES NEURAIS ARTIFICIAIS



FONTE: RAUBER, 1998

Durante o processo de aprendizagem os pesos normalmente percorrem uma modificação iterativa, onde o algoritmo julga a qualidade do peso e determina se deve sofrer uma modificação no valor de uma diferença na próxima iteração. A

aprendizagem pela Regra de Hebb, onde o peso de ligação entre dois neurônios que estão ativos ao mesmo tempo deve ser reforçado, a definição dessa regra baseia-se em estudos biológicos do cérebro. Esta regra define um algoritmo de adaptação dos pesos, sem definição de um objetivo a atingir.

Na aprendizagem pela Regra de Delta, é calculado na saída uma função, já sabendo o valor desejado, o objetivo do algoritmo é minimizar o erro entre os valores calculados pela rede e desejados pelos exemplos. Já na Aprendizagem Competitiva, considera-se redes de neurônios onde um único neurônio pode ser ativo ao mesmo tempo, ou seja, todos os outros possuem ativação igual a zero e somente o vencedor emite um sinal de ativação.

Dentro da aprendizagem supervisionada, utilizando a regra Delta, foi selecionado o Modelo *Perceptron* que foi desenvolvido nas décadas de 1950 e 1960 pelo cientista Frank Rosenblatt, sendo o mais comum usar outros modelos de neurônios artificiais, permitindo uma compreensão clara de como funciona uma rede neural em termos matemáticos. É um classificador que usa retro propagação para aprender um perceptron de várias camadas para classificar instâncias. A rede pode ser construída manualmente ou configurada usando uma heurística simples. Os parâmetros de rede também podem ser monitorados e modificados durante o tempo de treinamento. Os nós nesta rede são todos sigmóides (exceto quando a classe é numérica; nesse caso, os nós de saída se tornam unidades lineares sem limite).

3.5.3 MÁQUINA DE VETORES DE SUPORTE (SVM)

A metodologia de Máquina de Vetores foi desenvolvida por Vapnik (1995), com o intuito de resolver problemas de classificação de padrões. É uma técnica de *machine learning* supervisionado, recebe um conjunto de amostras na forma par ordenado (X, Y), e com base nessas informações será escolhida uma função, ou classificador, prevendo o valor esperado da amostra que é desconhecido. Essa implementação substitui globalmente todos os valores ausentes e transforma atributos nominais em binários. Também normaliza todos os atributos por padrão. Segundo HAYKIN (1999) a máquina de vetores suporte é uma outra categoria das redes neurais alimentadas adiante, ou seja, redes cujas saídas dos neurônios de uma camada alimentam os neurônios da camada posterior, não ocorrendo a realimentação.

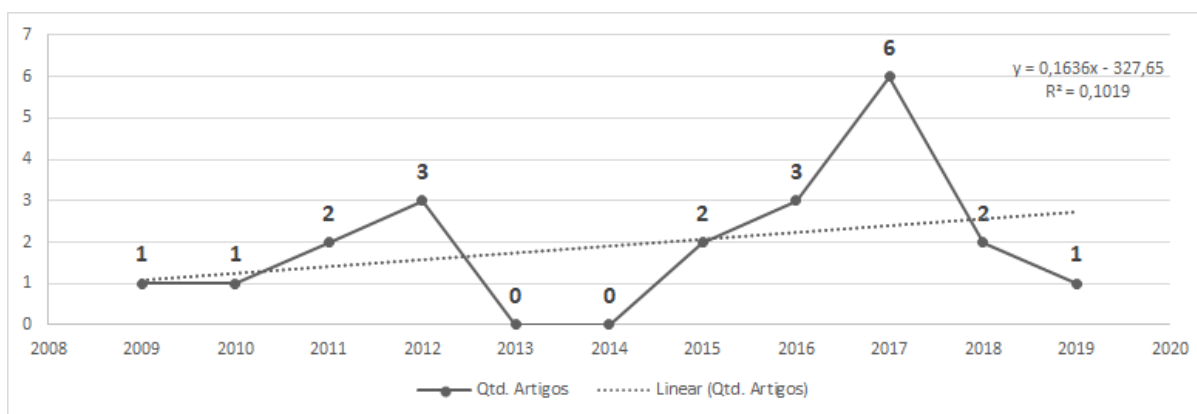
Algumas das principais características de SVM incluem: a boa capacidade de generalização, onde os classificadores gerados por uma SVM alcançam bons resultados de generalização, que é medida por sua eficiência na classificação de dados que não pertençam ao conjunto utilizado em seu treinamento; Robustez em grandes dimensões, diante de objetos de grandes dimensões, comumente há a ocorrência de overfitting nos classificadores; Convexidade da função objetivo, sua aplicação implica na otimização de uma função quadrática, que possui apenas um mínimo global; Teoria bem definida, onde possuem uma base teórica bem estabelecida na matemática e estatística.

4. ANÁLISE E DISCUSSÃO DOS DOCUMENTOS RECUPERADOS

4.1 RESULTADOS DAS ANÁLISES QUANTITATIVAS

Ao analisar a quantidade de publicações por ano e sua tendência, nota-se a concentração de publicações a partir de 2015, com cerca de 63,64% dos artigos publicados (14/22). O que confirma o aumento de interesse sobre o tema de análise preditiva de *churn* em *e-commerces*. (figura 13)

FIGURA 13: RELAÇÃO DE NÚMERO DE PUBLICAÇÕES COM LINHA DE TENDÊNCIA



FONTE: A AUTORA

Em relação à tipologia dos 22 registros recuperados, 59,09% dos registros consistem em artigos de periódicos (13/22) e 36,3636% correspondem a trabalhos publicados em eventos (8/22). Todos os registros recuperados foram elaborados em inglês, consistindo no único idioma retornado. Com relação às fontes de publicação, 17 fontes foram identificadas, destacando-se: a) com 3 registros, *Expert Systems With Applications*; que é um periódico internacional com foco em informações relacionadas a sistemas inteligentes e especialistas aplicados na indústria, governo e universidades em todo o mundo; b) com dois registros, *Information Sciences* que é uma revista projetada para atender pesquisadores, desenvolvedores, gerentes, planejadores estratégicos, estudantes de graduação e outros interessados em atividades de pesquisa de ponta em informações, engenharia do conhecimento e sistemas inteligentes; com dois registros; e c) com dois documentos *IEE Intelligence Systems*, que é uma publicação bimestral da *IEEE Computer Society*. Constatou-se ao aplicar a lei de Bradford que, mesmo que

com baixa quantidade de registros, a distribuição das fontes em zonas corresponde à suposição de que a primeira zona é composta por um pequeno grupo de periódicos mais relacionados ao tema; enquanto que a segunda e a terceira zonas exigem um número maior de periódicos conforme apresentado na tabela 3.

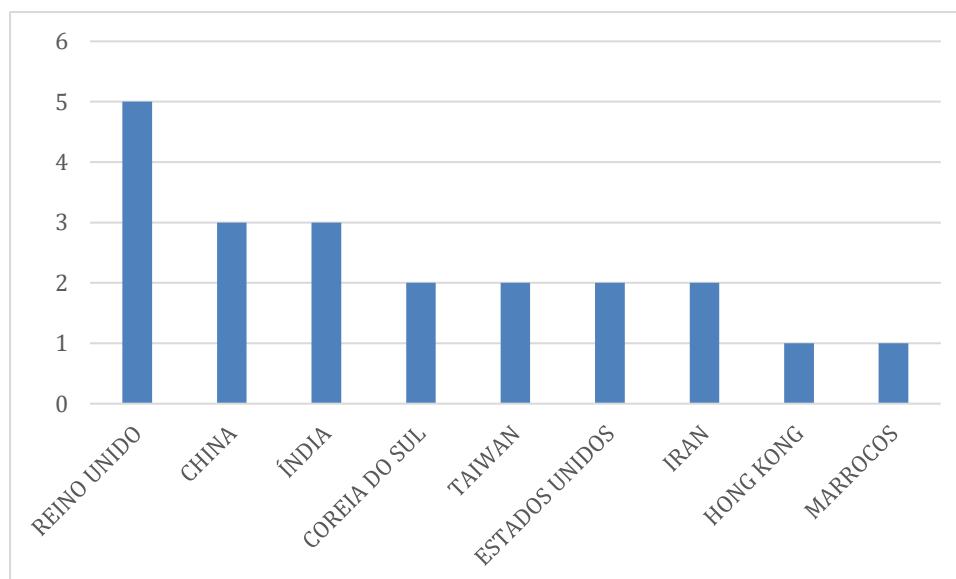
TABELA 3 - LEI DE BRADFORD.

	Quantidade de Periódicos	Total da Produção (%)
Zona 1	8	38,1%
Zona 2	7	33,3%
Zona 3	7	33,3%

FONTE: A AUTORA (2019)

Quanto ao país de origem das publicações dos documentos, apenas 9 países possuem publicações, destacando-se o Reino Unido, China e Índia (figura 14).

FIGURA 14: PAÍSES ORIGEM DAS PUBLICAÇÕES



FONTE: A AUTORA (2019)

Ao traduzir os títulos, palavras-chave plus, palavras-chave e títulos dos documentos e inseri-los em uma nuvem de palavras, obtém-se: Quanto aos títulos dos documentos (à esquerda): os maiores destaques são os tópicos "Previsão", "Churn" e "clientes"; e "Dados" e "Padrões", ambos com seis aparições. Em relação às Palavras-chave plus dos documentos, ou seja, aquelas acrescentadas pela

própria base de dados (à direita), os principais tópicos consistem em "Vendas", "Previsão" e "Vendas", conforme apresentados na figura 15.

Palavras-chave são palavras e frases importantes de um trabalho de pesquisa, expressando sua essência. As palavras-chave plus são definidas pela própria Scopus, enquanto as palavras-chaves e títulos são definidos pelos próprios autores.

FIGURA 15: NUVEM DE PALAVRAS: TÍTULOS E PALAVRAS-CHAVES PLUS



FONTE: A AUTORA (2019)

A tabela 4 apresenta a frequência das palavras-chaves plus e o número de vezes que foram utilizadas. Dos 208 termos, 80,17% (173/208) ocorre apenas uma vez. A palavra-chave mais utilizado é “Vendas”, com frequência 14; seguida de “Previsão”, com frequência 12 e; “Mineração de Dados” e “Árvores de Decisão” com frequência 7.

TABELA 4 - FREQUÊNCIA DAS PALAVRAS-CHAVE PLUS

# Palavras-Chave	Freq.	Descrição
1	14	Vendas
1	12	Previsão
2	7	Mineração De Dados; Árvores De Decisão
1	6	Comércio
2	5	Comércio Eletrônico; Previsões De Churn
4	4	Previsão De Churn De Clientes; Sistemas De Aprendizagem; Comércio Eletrônico; Redes Neurais
4	3	Churn De Clientes; Satisfação Do Cliente; Regressões Logísticas; Análise De Regressão
20	2	Técnicas De Mineração De Dados; Suporte A Máquinas Vetoriais; Algoritmos De Aprendizagem; Previsão De Comportamento; Inteligência Artificial; Rede Neural Artificial; Dados Desequilibrados; Redes Sociais (Online; Testes Estatísticos; Conjuntos De Dados; Modelo De Previsão; Estudos Empíricos; Relações Públicas; Lucratividade; Data Mining; Algoritmos; E-Commerce; Marketing; Educação; Crime
173	1	...
208		

FONTE: A AUTORA (2019)

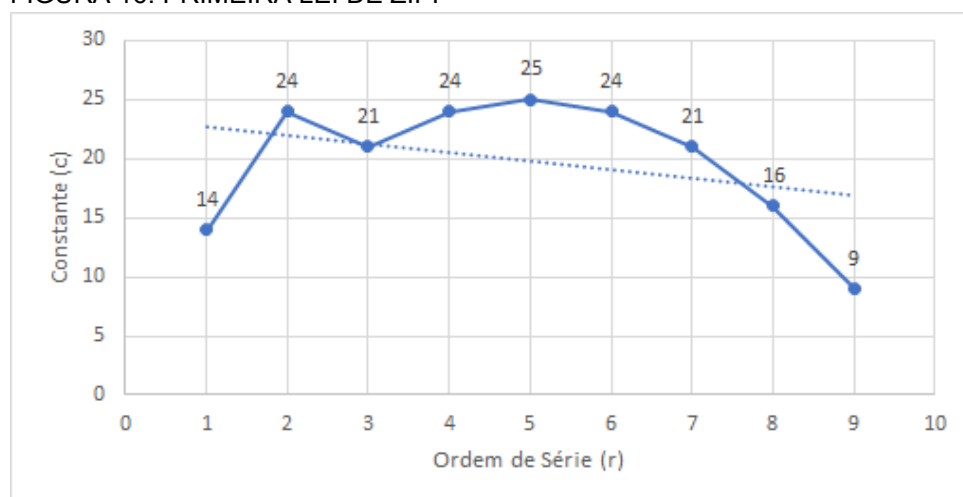
Aplicando a primeira Lei de Zipf, onde explica que há palavras que repetem muito mais do que as outras, e essas obtêm uma alta frequência. Ordenando as palavras-chave decrescentemente de acordo com a sua frequência de ocorrência, obteve-se uma ordem de série "r". Para obter a constante da primeira lei de Zipf ("c"), multiplica-se a ordem da série ("r") e a frequência da ocorrência ("f"), conforme apresentado na tabela. Observa-se na figura 16, o ponto máximo que representa o produto da ordem de série cinco. Nota-se uma parábola na qual a linha tracejada demonstra a tendência decrescente resultante da série multiplicada pela frequência de ocorrência.

TABELA 5 - FREQUÊNCIA DAS PALAVRAS-CHAVE PARA APLICAÇÃO DA LEI DE ZIPF

Qtd de palavras-chave	Ordem de Série "r"	Frequência de Ocorrência "f"	Constante da 1ª Lei de Zipf (r.f=c)
1	1	14	14
1	2	12	24
2	3	7	21
1	4	6	24
2	5	5	25
4	6	4	24
4	7	3	21
20	8	2	16
173	9	1	9

FONTE: A AUTORA (2019)

FIGURA 16: PRIMEIRA LEI DE ZIPF



FONTE: A AUTORA (2019)

A figura 17 mostra a nuvem de palavras referentes às palavras-chave dos autores (esquerda) e o resumo apresentado pelos mesmos (direita). As palavras-chave mais utilizadas são "Churn", "Clientes", "Previsão" e "E-commerce"; bem como "Dados" e "Aprendizagem". Nos resumos as palavras com maiores destaques são: "Clientes", "Churn", "Previsão", "Modelo" e "Recursos".

FIGURA 17: PALAVRAS-CHAVE DOS AUTORES E RESUMO



FONTE: A AUTORA (2019)

4.2 RESULTADOS DAS ANÁLISES DE CONTEÚDO

A respeito dos documentos mais citados, os cinco registros com o maior número de citações são elencados na Tabela 6.

TABELA 6 - ARTIGOS MAIS CITADOS

Título do Artigo	Total de Citações	Ano de Publicação	Citações por Ano
A HIERARCHICAL MULTIPLE KERNEL SUPPORT VECTOR MACHINE FOR CUSTOMER CHURN PREDICTION USING LONGITUDINAL BEHAVIORAL DATA	56	2012	8,0
CRM STRATEGIES FOR A SMALL-SIZED ONLINE SHOPPING MALL BASED ON ASSOCIATION RULES AND SEQUENTIAL PATTERNS	42	2012	6,0
AN EXTENDED SUPPORT VECTOR MACHINE FORECASTING FRAMEWORK FOR CUSTOMER CHURN IN E-COMMERCE	34	2011	4,3
AN EMPIRICAL COMPARISON OF TECHNIQUES FOR THE CLASS IMBALANCE PROBLEM IN CHURN PREDICTION	15	2017	7,5
DEVELOPING A PREDICTION MODEL FOR CUSTOMER CHURN FROM ELECTRONIC BANKING SERVICES USING DATA MINING	9	2016	3,0

FONTE: A AUTORA (2019)

O primeiro artigo “*A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data*” (CHEN, Zhen-Yu; FAN, Zhi-Ping; SUN, Minghe, 2012) utilizou Não apenas dados estáticos para o modelo preditivo, mas também dados longitudinais para melhorar o desempenho na previsão de *churn* de clientes. Normalmente, os dados comportamentais

longitudinais são transformados em dados estáticos antes de serem incluídos em um modelo de previsão. Neste estudo, uma estrutura com técnicas de conjunto é apresentada para a previsão de *churn* de clientes diretamente usando dados comportamentais longitudinais. Utiliza uma abordagem chamada de máquina vetorial de suporte de múltiplos núcleos hierárquicos (H-MK-SVM). Um algoritmo de treinamento de três fases para o H-MK-SVM é desenvolvido, implementado e testado. O H-MK-SVM constrói uma função classification estimando o coefficients de ambas as variáveis comportamentais estáticas e longitudinais no processo de treinamento sem transformação dos dados comportamentais longitudinais. O processo de treinamento do H-MK-SVM também é um processo de seleção de características e sequências de tempo porque o Não -zero esparso coefficients corresponde à variável selecionada. Experimentos computacionais usando três bases de dados do mundo real foram conduzidos. Os resultados computacionais usando múltiplos critérios de medição de desempenho mostram que o H-MK-SVM diretamente usando dados comportamentais longitudinais têm um desempenho melhor do que os classificadores atualmente disponíveis.

O segundo documento “*CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns*” (Shim B;Choi K;Suh Y, 2012) Propõe estratégias de CRM para shopping centers de pequeno porte baseados em regras de associação e padrões sequenciais obtidos pela análise dos dados de transação da loja. Primeiro foi realizada a classificação dos clientes como VIPs e Não -VIPs. Depois, foi buscado regras de associação e padrões sequenciais entre as categorias de itens, bem como entre as subcategorias de itens da transação dados de apenas clientes classificados como VIPs pela classificação modelo. Utilizando Árvore de decisão, Rede neural artificial, Regressão logística e "Bagging" com cada um deles como classificador base. DT, ANN e LR são abordagens de mineração de dados que foram muito usadas para classificação e / ou previsão no suporte à decisão e demonstraram bom desempenho (Para classificação de VIP e Não VIP).

O documento “*An extended support vector machine forecasting framework for customer churn in e-commerce*” (Yu X;Guo S;Guo J;Huang X, 2011) estabelece um quadro de previsão de *churn* de clientes em quatro etapas. Primeiro, coleta os dados de comportamento do cliente e os converte no data warehouse (DW) por meio de uma Extract, Transform, Load (ETL). Depois, extrai algumas amostras

como objetos de treino. Como terceiro passo, é escolhido os algoritmos de predição para treinar amostras selecionadas. Por último o algoritmo de predição é selecionado para prever outros clientes, sendo eles: Rede neural artificial (ANN), árvore de decisão, SVM e ESVM. O resultado mostra que o ESVM tem melhor desempenho entre eles no aspecto de precisão, taxa de acerto, taxa de cobertura, coeficiente de elevação e tempo de tratamento.

“*An empirical comparison of techniques for the class imbalance problem in churn prediction*” (Zhu B;Baesens B;Vanden Broucke Sklm, 2017), teve como objetivo comparar métodos avançados para lidar com desequilíbrios de classe em conjuntos de dados de previsão de *churn* no mundo real. Foi utilizado um critério de lucro máximo esperado, chamada Medida de lucro esperado (EMP), e é usado como uma das principais medidas de desempenho para oferecer mais insights do ponto de vista do custo-benefício. Os resultados experimentais mostram que a métrica de avaliação aplicada tem um grande impacto no desempenho das técnicas.

“Developing a prediction model for customer *churn* from electronic banking services using data mining” (Keramati A;Ghaneei H;Mirmohammadi Sm, 2016), identificou fatores que o *churn* de usuários de serviços de bancos eletrônicos têm, usando dados demográficos, das transações realizadas através de portais bancários, o tempo de duração da associação do consumidor e as reclamações do cliente, usando CRISP-DM, onde seguiu seis passos: Entender o negócio, entender os dados, pré-processar os dados, modelar, avaliar e deploy.

A figura 18 apresenta a nuvem de palavras que relaciona os principais métodos de *machine learning* utilizados. Destacando-se Árvore de Decisão e Rede Neural Artificial (RNA) e Máquina de Vetores de Suporte (SVM) com seis aparições cada; e Regressão Logística e Floresta Aleatória (*Random Forest*) com quatro aparições.

FIGURA 18: MÉTODOS UTILIZADOS



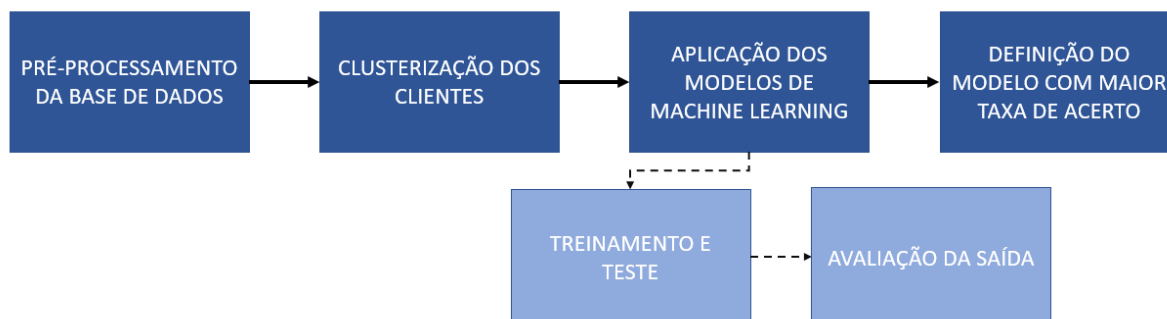
FONTE: A AUTORA (2019)

Quanto à metodologia utilizada 42,86% (9/28) dos documentos analisados realizaram primeiro a classificação de uma base de clientes pré-existente, a classificação dividia os clientes entre clientes VIPs e não VIPs; 38,10% (8/21) dos documentos comparam três ou mais métodos de *machine learning* para previsão de *churn*, apenas 37,50% (3/8) explicita a base utilizada para análise; 14,28% (3/21) utilizaram Não apenas dados estáticos para o modelo preditivo, mas também dados longitudinais; 5,76% trata-se de uma revisão da literatura.

Das bases de dados utilizadas apenas 4 documentos divulgaram o nome das empresas ou comentam do que se trata as empresas que cederam a base de dados para a análise, porém nenhuma utiliza uma base de dados pública. Dentre eles: E-commerce Coreano de pequeno porte, utilizando Tabela demográfica, Tabela de boletins, Tabela de comentários, Tabela de gerenciamento de pedidos e tabela de produtos; E-Commerce ASOS, utilizando dados demográficos dos clientes, histórico de compras, histórico de devoluções, registros de sessão da web e do aplicativo; Site de Namoro.

De modo geral, os documentos que apresentaram a utilização de alguma base de dados seguiram a mesma metodologia, conforme apresentada na figura 19.

FIGURA 19: SÍNTESE DA METODOLOGIA UTILIZADA



FONTE: A AUTORA (2019)

Primeiro realizavam o pré-processamento da base de dados, retirados normalmente de um datawarehouse (DW), depois seguiam para a classificação dos clientes, clusterizando-os normalmente em dois grupos (VIPs e não VIPs), e por último buscavam regras de associação e padrões sequenciais aplicando alguns modelos de *machine learning* para predição.

5. APLICAÇÃO DOS MODELOS DE MACHINE LEARNING

Para o teste dos modelos de *machine learning* selecionados foram utilizadas 155.677 instâncias com 9 atributos, conforme apresentado na figura 20. Em todos os modelos foi usado a opção de teste em *cross-validation*, com 10 *folds* onde é dividido o *dataset* em 10 partes, usando 10-1 partes para treino e a parte remanescente para teste, fazendo isso 10 vezes. Em cada uma das 10 vezes, testa-se o modelo com um *fold* diferente calculando a métrica escolhida para avaliação do modelo.

FIGURA 20: INFORMAÇÕES PROCESSADAS

```

Relation:    base_cliente54
Instances:   155677
Attributes:  9
             i»¿tipopessoa
             estado
             newsletter
             valorgasto
             pedidos
             taxaaprovacao
             cadastroprimeiracompra
             primeiraultimacompra
             churn
Test mode:   10-fold cross-validation

```

FONTE: A AUTORA (2019)

Os dados foram discretizados, onde valores numéricos foram transformados em nominais ou discretos para que possam representar melhor os dados no conjunto. A tabela 7 representa a discretização e os atributos discretizados.

TABELA 7 - DADOS DISCRETIZADOS

taxaaprova cao			primeiraultimacompra		
Min.	Máx.	Discretização	Min.	Máx.	Discretização
0	0.33	A	-70	R\$ 51	A
0.33	0.66	B	51	R\$172	B
0.66	1.00	C	172	R\$293	C

Pedidos			cadastroprimeiracompra		
Min.	Máx.	Discretização	Min.	Máx.	Discretização
0	11	A	-70	R\$ 48	
11	21	B	48	R\$166	
21	31	C	166	R\$284	

valorgasto		
Min.	Máx.	Discretização
0	R\$18,430	A
18430	R\$36,860	B
36860	R\$55,290	C

FONTE: A AUTORA (2019)

A discretização dos atributos foi realizada em grupos de três, onde foram encontrados intervalos com valor máximo e mínimo.

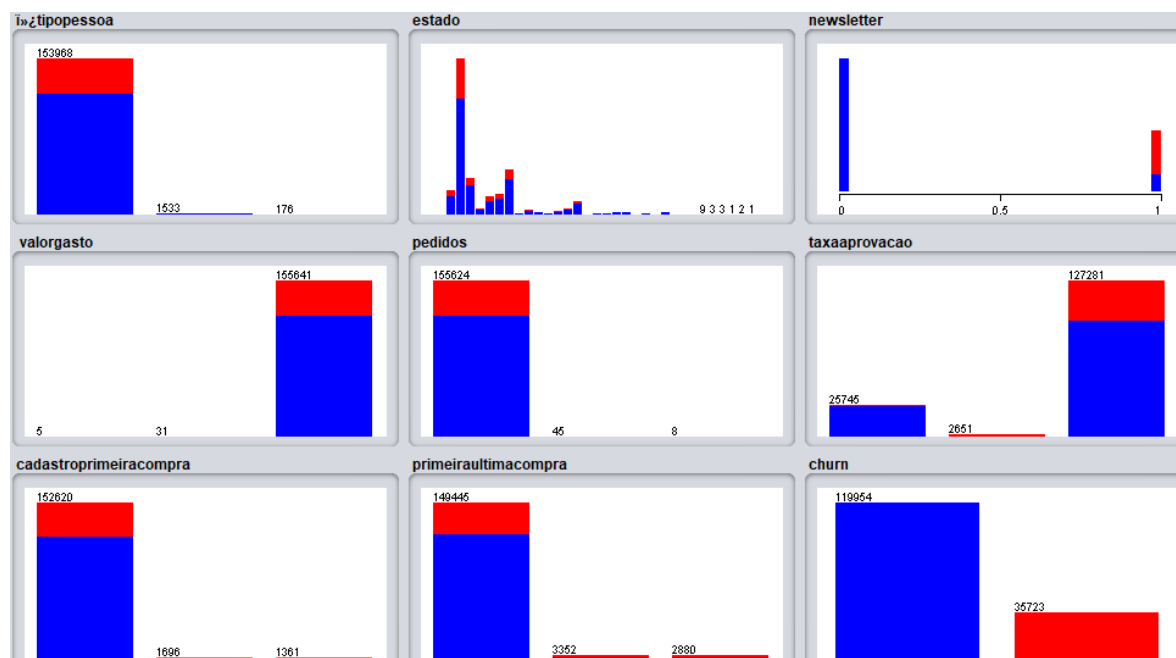
QUADRO 1 - DESCRIÇÃO DA BASE DE DADOS APÓS DISCRETIZAÇÃO

Nome do Atributo	Tipo	Descrição	Valor
idclcli	Númerica	ID de cada cliente	-
tipopessoa	Nominal	Pessoa Física ou Jurídica	F; J
estado	Nominal	Estado em que o cliente mora	PR; SP; MG; ES; SC; RS; RJ; PI; GO; CE; SE; DF; PE; BA; AM; AL; MT; MS; PB; TO; MA; PA; RN; RO; AC; RR; AP
newsletter	Binário	Cliente aceitou receber emails com promoções	1;0
valorgasto	Númerica	Soma do valor absoluto comprado pelo cliente	A;B;C
pedidos	Númerica	Quantidade de pedidos realizados pelo cliente	A;B;C
taxaaprova cao	Númerica	Taxa de aprovação do cliente (valor realmente pago/valor comprado)	A;B;C
cadastroprimeiracompra	Númerica	Diferença de dias entre o cadastro e a primeira compra	A;B;C
primeiraultimacompra	Númerica	Diferença de dias entre a primeira compra e a última compra	A;B;C
churn	Nominal	Definição do cliente entre chun (Sim) ou não Chum (Não)	SIM; NÃO

FONTE: A AUTORA (2019)

Uma visualização gráfica em formato de histograma mostrando cada atributo da base de dados e seu domínio foi gerado pelo Weka é apresentado na figura 21.

FIGURA 21: REPRESENTAÇÃO GRÁFICA DOS ATRIBUTOS GERADA PELO WEKA



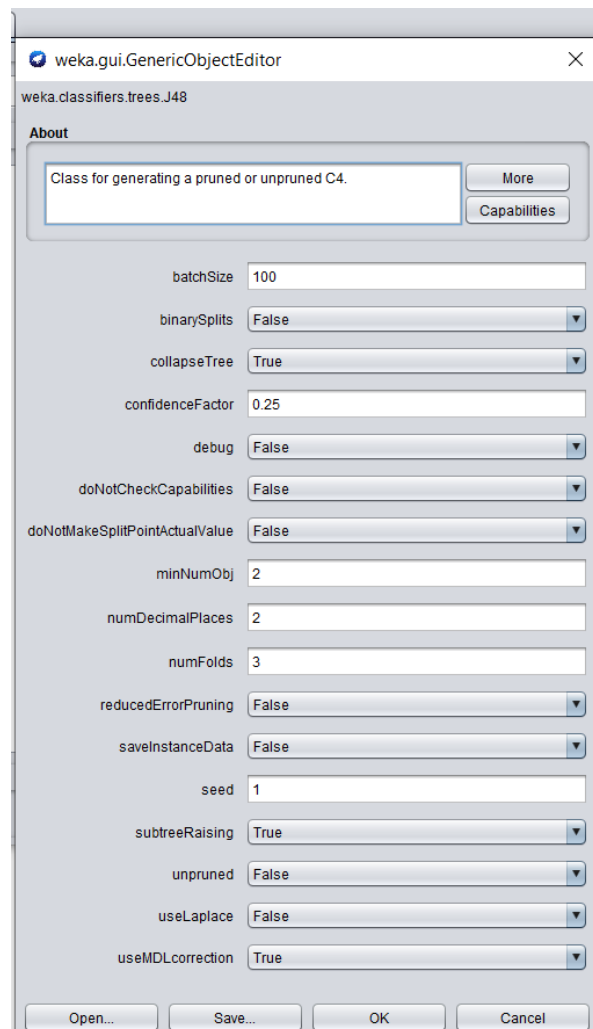
FONTE: A AUTORA (2019)

A cor azul representa os clientes *Churn* e a vermelha os clientes *Não-Churn*. A base de dados é formada por 119.954 (77,05%) churners e 35.723 (22,95%) não churners

5.1.1 ÁRVORE DE DECISÃO (J48)

O algoritmo J48 é uma implementação da Árvore de Decisão C4.5. A figura 22 apresenta os parâmetros default utilizados no modelo.

FIGURA 22: PARÂMETROS DEFAULT J48



FONTE: DADOS DA PESQUISA UTILIZANDO O SOFTWARE WEKA (2019)

O Weka apresenta um resumo sobre o significado das opções dos parâmetros:

- *seed*: Usada para aleatorizar os dados quando se utiliza a poda de erro reduzido;
- *unpruned*: Se a poda é feita;
- *confidenceFactor*: O fator de confiança usado na poda (valores menores incorrem em mais podas);
- *numFolds*: Determina a quantidade de dados usados na poda com erros reduzidos. Uma dobra é usada para a poda, o resto para o crescimento da árvore;
- *numDecimalPlaces*: O número de casas decimais a ser usado para a saída de números no modelo;

- *batchSize*: O número preferido de instâncias para processar se a previsão de lote estiver sendo executada. Mais ou menos instâncias podem ser fornecidas, mas isso dá às implementações a chance de especificar um tamanho de lote preferido;
- *reducedErrorpruning*: Se a poda com erro reduzido é usada em vez da poda C.4.5;
- *useLaplace*: Se as contagens nas folhas são suavizadas com base no Laplace;
- *doNotMakeSplitPointActualValue*: Se for *true*, o ponto de divisão Não é realocado para um valor de dados real. Isso pode render acelerações substanciais para grandes conjuntos de dados com atributos numéricos;
- *debug*: Se definido como *true*, o classificador pode gerar informações adicionais para o console;
- *subtreeRaising* : Se deve considerar a operação de elevação da sub-árvore durante a poda;
- *saveInstanceData*: Se deve guardar os dados de treino para visualização.
- *binarySplits*: Se usar divisórias binárias em atributos nominais ao construir as árvores;
- *doNotCheckCapabilities*: Se configurado, as capacidades do classificador não são verificadas antes que o classificador seja compilado (Use com cuidado para reduzir o tempo de execução);
- *minNumObj*: O número mínimo de instâncias por folha;
- *useMDLcorrection*: Se a correção MDL é usada ao encontrar divisões em atributos numéricos;
- *collapseTree*: Se as partes são removidas que não reduzem o erro de treinamento.

Os resultados são apresentados na figura 23. O resultado encontra-se disposto no Apêndice B.

FIGURA 23: RESULTADOS DO J48 NO WEKA

```

Number of Leaves :      26

Size of the tree :      39

Time taken to build model: 1.31 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances   147093           94.486 %
Incorrectly Classified Instances    8584           5.514 %
Kappa statistic                  0.8548
Mean absolute error              0.0898
Root mean squared error          0.2121
Relative absolute error          25.4005 %
Root relative squared error       50.4305 %
Total Number of Instances        155677

```

FONTE: A AUTORA (2019).

Foram classificados corretamente 94,48% das amostras. O Erro Absoluto médio, expressa o erro em uma maneira de fácil compreensão. Medida que indica a média do afastamento de todos os valores fornecidos pelos classificadores e o seu real valor e foi de 0,0898; A Raiz quadrada do erro quadrático médio, expressa a extensão da variação dos dados foi de 0,2121. A figura 24 apresenta diversas métricas de desempenho por classe avaliada, com estas informações é possível avaliar Falso Positivo, Falso Negativo, Precisão, métricas relacionadas a Curva ROC geral do modelo ou por classes

FIGURA 24: MÉTRICAS DE DESEMPENHO POR CLASSE AVALIADA

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,933	0,015	0,995	0,933	0,963	0,862	0,966	0,990	Sim
	0,985	0,067	0,814	0,985	0,891	0,862	0,966	0,820	NÃO
Weighted Avg.	0,945	0,027	0,954	0,945	0,947	0,862	0,966	0,951	

FONTE: A AUTORA (2019)

Os valores obtidos que indicam que é possível identificar que o padrão detectado acertar mais vezes em um dos casos, é apresentado no *True Positive (TP)* onde a classe não possui mais acertos, com 98,5% do que em Sim, que apresentou o menor TP representado por 93,30%. Os resultados obtidos indicam

os dados classificados errados como positivos (False Positive (FP)), sendo a maior probabilidade encontrada em Sim, com 15% e a menor em não, com 6,7%. A precisão indica o valor do número de casos positivos por total de casos cobertos, os resultados foram melhores para a classificação de Sim apresentando 99,50% de precisão, enquanto não apresentou 81,40%. O resultado de cobertura (recall) é equivalente aos valores obtidos no verdadeiro positivo (TP), considerando que corresponde ao valor da cobertura de casos. Os valores obtidos no *F-Measure* indicam um desempenho de 96,30% para a classificação de Sim e de 89,1% para a classificação de não. Já os resultados obtidos em MCC demonstram um desempenho de 86,20% para ambas as classes. O *ROC Area* dos classificadores ficou em 93,60%, representando um bom resultado, considerando que um classificador excelente obtém valores próximos a 1. O *PRC Area* do classificador Sim ficou em 99,00%, enquanto não foi 82,00%

A figura 25 apresenta a matriz de confusão, sendo possível notar que 526 amostras NÃO foi predita como SIM (coluna a), e 8.058 SIM preditos como NÃO.

FIGURA 25: MATRIZ DE CONFUSÃO

```

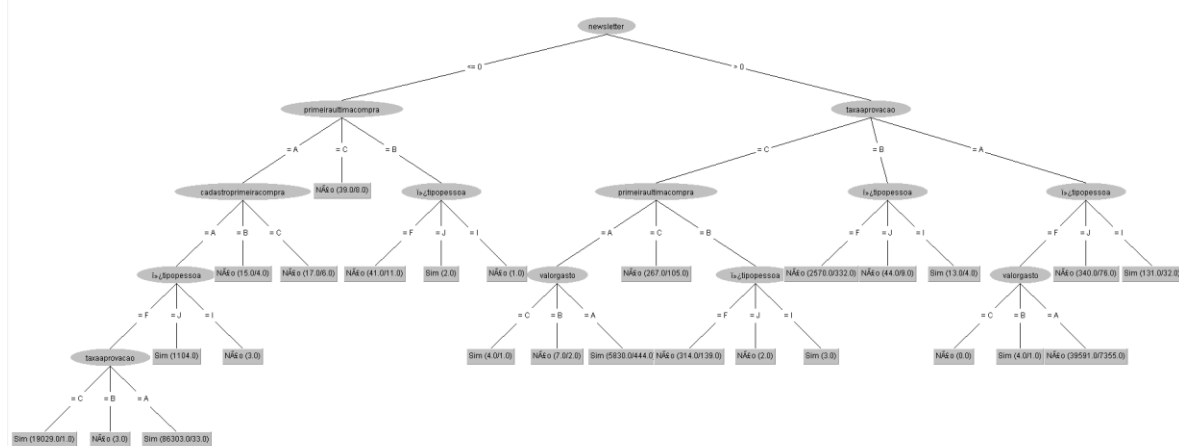
=== Confusion Matrix ===
      a      b  <-- classified as
111896  8058 |      a = Sim
  526 35197 |      b = NÃO

```

FONTE: A AUTORA (2019)

A figura 26 é uma representação gráfica da árvore de decisão criada. Ao executar o algoritmo J48 com os parâmetros default foi gerada uma árvore com 69 folhas e tamanho 94 em 3,4 segundos. Porém devido ao seu tamanho, as folhas ficaram sobrepostas, impossibilitando a análise completa. No apêndice B é apresentada a árvore J48 descrita com o resultado completo.

FIGURA 26: REPRESENTAÇÃO GRÁFICA DA ÁRVORE J48



FONTE: A AUTORA (2019)

Na árvore retirada da ferramenta Weka é possível identificar que a raiz da árvore corresponde ao *Newletter*, sendo este, portanto, o atributo com maior influência. O segundo atributo com maior influência quando o cliente não aceita receber e-mails é a a diferença de dias entre a primeira e última compra. Quando o cliente aceita receber e-mails, o atributo com maior influência é a taxa de aprovação, diferença de primeira e ultima compra, depois o tipo de pessoa. A seguir, a árvore textual:

newsletter <= 0

```

| primeiraultimacompra = A
| | cadastroprimeiracompra = A
| | | tipopessoa = F
| | | | taxaaprovacao = C: Sim (19029.0/1.0)
| | | | taxaaprovacao = B: Não (3.0)
| | | | taxaaprovacao = A: Sim (86303.0/33.0)
| | | tipopessoa = J: Sim (1104.0)
| | | tipopessoa = I: Não (3.0)
| | cadastroprimeiracompra = B: Não (15.0/4.0)
| | cadastroprimeiracompra = C: Não (17.0/6.0)
| primeiraultimacompra = C: Não (39.0/8.0)
| primeiraultimacompra = B
| | tipopessoa = F: Não (41.0/11.0)
| | tipopessoa = J: Sim (2.0)
| | tipopessoa = I: Não (1.0)

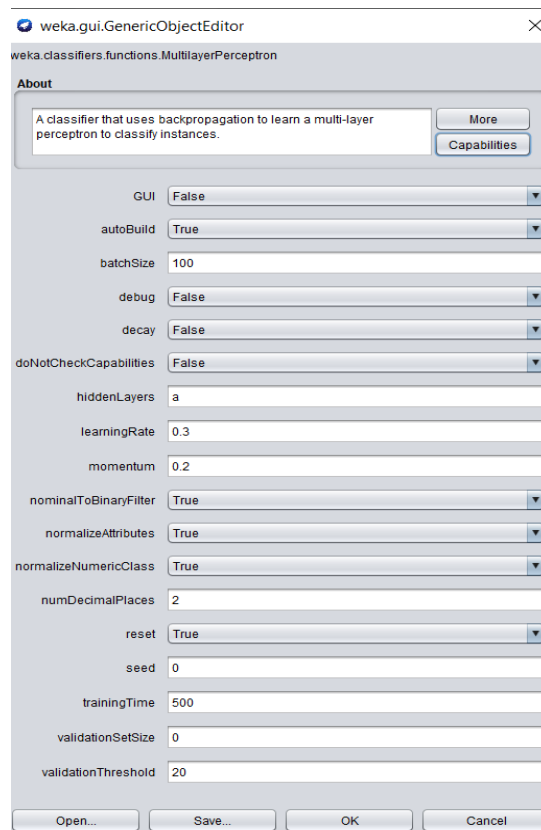
```

newsletter > 0

```
| taxaaprovacao = C
| | primeiraultimacompra = A
| | | valorgasto = C: Sim (4.0/1.0)
| | | valorgasto = B: Não (7.0/2.0)
| | | valorgasto = A: Sim (5830.0/444.0)
| | primeiraultimacompra = C: Não (267.0/105.0)
| | primeiraultimacompra = B
| | | tipopessoa = F: Não (314.0/139.0)
| | | tipopessoa = J: Não (2.0)
| | | tipopessoa = I: Sim (3.0)
| taxaaprovacao = B
| | tipopessoa = F: Não (2570.0/332.0)
| | tipopessoa = J: Não (44.0/9.0)
| | tipopessoa = I: Sim (13.0/4.0)
| taxaaprovacao = A
| | tipopessoa = F
| | | valorgasto = C: Não (0.0)
| | | valorgasto = B: Sim (4.0/1.0)
| | | valorgasto = A: Não (39591.0/7355.0)
| | tipopessoa = J: Não (340.0/76.0)
| | tipopessoa = I: Sim (131.0/32.0)
```

5.1.2 REDE NEURAL ARTIFICIAL (RNA)

O weka define *Multilayer Perceptron* como um classificador que usa *backpropagation* para aprender um *perceptron* multicamadas para classificar instâncias. A rede pode ser construída manualmente ou configurada usando uma heurística simples. Os parâmetros da rede também podem ser monitorados e modificados durante o tempo de treinamento. Os nós desta rede são todos sigmóides. A figura 27 apresenta os parâmetros default utilizados no modelo.

FIGURA 27: PARÂMETROS DEFAULT *MULTILAYER PERCEPTRON*

FONTE: DADOS DA PESQUISA UTILIZANDO O SOFTWARE WEKA (2019)

O Weka apresenta um resumo sobre o significado das opções dos parâmetros:

- *seed*: Semente usada para inicializar o gerador de números aleatórios. Números aleatórios são usados para definir os pesos iniciais das conexões entre nós, e também para embaralhar os dados de treinamento;
- *momentum*: Momento aplicado às atualizações de peso;
- *nominalToBinaryFilter*: Isso pré-processa as instâncias com o filtro *NominalToBinary*. Isso pode ajudar a melhorar o desempenho se houver atributos nominais nos dados;
- *hiddenLayers*: Define as camadas ocultas da rede neural. Esta é uma lista de números inteiros positivos. 1 para cada camada oculta. Separação por vírgula. Para Não ter camadas ocultas, coloque um único 0 aqui. Esta definição de camada só será usada se a auto-

construção estiver definida. Há também valores wildcard: 'a' = (attribs + classes) / 2, 'i' = attribs, 'o' = classes, 't' = attribs + classes;

- *validationThreshold*: Usado para terminar o teste de validação. O valor aqui dita quantas vezes em uma linha o erro do conjunto de validação pode piorar antes do treinamento ser terminado;
- GUI: Abre uma interface gui. Isso permitirá a pausa e alteração da rede neural durante o treinamento;
- *normalizeAttributes*: Isso irá normalizar os atributos. Isso pode ajudar a melhorar o desempenho da rede. Isso Não depende da classe ser numérica. Isto também irá normalizar os atributos nominais (depois de terem sido executados através do filtro nominal para binário se estiver em uso) para que os valores binários estejam entre -1 e 1;
- *numDecimalPlaces*: O número de casas decimais a ser usado para a saída de números no modelo;
- *batchSize*: O número preferido de instâncias a serem processadas se a previsão de lote estiver sendo executada. Mais ou menos instâncias podem ser fornecidas, mas isso dá às implementações a chance de especificar um tamanho de lote preferido;
- *decay*: Isso fará com que a taxa de aprendizagem diminua. Isso dividirá a taxa de aprendizagem inicial pelo número da época para determinar qual deve ser a taxa de aprendizagem atual. Isso pode ajudar a impedir que a rede se afaste do output alvo, bem como melhorar o desempenho geral. Note que a taxa de aprendizagem decadente não será mostrada na GUI, apenas a taxa de aprendizagem original. Se a taxa de aprendizagem for alterada na GUI, esta será tratada como a taxa de aprendizagem inicial;
- *validationSetSize*: O tamanho percentual do conjunto de validação. (O treinamento continuará até que seja observado que o erro no conjunto de validação foi consistentemente piorando, ou se o tempo de treinamento for atingido);
- *trainingTime*: O número de épocas a treinar. Se o conjunto de validação for diferente de zero, então pode terminar a rede mais cedo;
- debug: Se definido como *true*, o *classificador* pode gerar informações adicionais para o console;

- *autoBuild*: Adiciona e conecta camadas ocultas na rede;
- *normalizeNumericClass*: Isso irá normalizar a classe se ela for numérica. Isso pode ajudar a melhorar o desempenho da rede. Ele normaliza a classe para estar entre -1 e 1. Note que isso é apenas internamente, a saída será reduzida para o intervalo original;
- *learningRate*: A taxa de aprendizagem para atualizações de peso;
- *doNotCheckCapabilities*: Se configurado, as capacidades do classificador não são verificadas antes que o classificador seja compilado (Use com cuidado para reduzir o tempo de execução);
- *reset*: Isso permitirá que a rede seja resetada com uma taxa de aprendizagem mais baixa. Se a rede divergir da resposta, isto irá automaticamente redefinir a rede com uma taxa de aprendizagem mais baixa e começar o treinamento novamente. Esta opção só está disponível se a GUI Não estiver definida.

Foi classificado corretamente 94,50% das amostras. O Erro Absoluto médio, expressa o erro em uma maneira de fácil compreensão. Medida que indica a média do afastamento de todos os valores fornecidos pelos classificadores e o seu real valor e foi de 0,083; A Raiz quadrada do erro quadrático médio, expressa a extensão da variação dos dados foi de 0,2069, conforme apresentado na figura 28.

FIGURA 28: RESULTADOS DO *MULTILAYER PERCEPTRON*

```

Time taken to build model: 1976.41 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      147066           94.4687 %
Incorrectly Classified Instances    8611             5.5313 %
Kappa statistic                    0.8542
Mean absolute error                 0.0892
Root mean squared error             0.2125
Relative absolute error             25.2337 %
Root relative squared error         50.5256 %
Total Number of Instances          155677

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,933	0,017	0,995	0,933	0,963	0,861	0,968	0,991	Sim
	0,983	0,067	0,815	0,983	0,891	0,861	0,968	0,829	Não
Weighted Avg.	0,945	0,029	0,953	0,945	0,946	0,861	0,968	0,954	

```

=== Confusion Matrix ===

      a      b  <-- classified as
111961  7993 |      a = Sim
   618 35105 |      b = Não

```

FONTE: A AUTORA (2019)

De modo geral os resultados apresentados na acurácia da classe foram bem parecidos com os resultados do J48.

5.1.3 MÁQUINA DE VETORES DE SUPORTE (SVM)

O Weka define SMO como algoritmo de otimização mínima sequencial de John Platt para treinamento de um classificador vetorial de suporte. Substitui globalmente todos os valores em falta e transforma atributos nominais em binários, normaliza todos os atributos por padrão. Problemas multiclasse são resolvidos usando a classificação em pares (aka 1-vs-1). Para obter estimativas de probabilidade adequadas, use a opção que ajusta os modelos de calibração aos outputs da máquina vetorial suporte. A figura 29 apresenta os parâmetros default utilizados no modelo.

FIGURA 29: PARÂMETROS DEFAULT SMO

weka.classifiers.functions.SMO

About

Implements John Platt's sequential minimal optimization algorithm for training a support vector classifier.

batchSize: 100

buildCalibrationModels: False

c: 1.0

calibrator: Choose Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

checksTurnedOff: False

debug: False

doNotCheckCapabilities: False

epsilon: 1.0E-12

filterType: Normalize training data

kernel: Choose PolyKernel -E 1.0 -C 250007

numDecimalPlaces: 2

numFolds: -1

randomSeed: 1

toleranceParameter: 0.001

Open... Save... OK Cancel

FONTE: DADOS DA PESQUISA UTILIZANDO O SOFTWARE WEKA (2019)

O Weka apresenta um resumo sobre o significado das opções dos parâmetros:

- *buildCalibrationModels*: Ajusta modelos de calibração aos outputs do SVM (para estimativas de probabilidade adequadas);
- *numFolds*: O número de dobras para validação cruzada usado para gerar dados de treinamento para modelos de calibração (-1 significa usar dados de treinamento);
- *randomSeed*: Semente de número aleatório para a validação cruzada;
- *c*: O parâmetro de complexidade C;
- *numDecimalPlaces*: O número de casas decimais a serem usadas para a saída de números no modelo;
- *batchSize*: O número preferido de instâncias a serem processadas se a previsão de lote estiver sendo executada. Mais ou menos instâncias

podem ser fornecidas, mas isso dá às implementações a chance de especificar um tamanho de lote preferido;

- *kernel*: O kernel a ser usado;
- *checksTurnedOff*: Desativa verificações demoradas;
- *debug*: Se definido como *true*, o *classificador* pode gerar informações adicionais para o console;
- *filterType*: Determina como/se os dados serão transformados;
- *toleranceParameter*: O parâmetro de tolerância (Não deve ser alterado);
- *calibrator*: O método de calibração a ser usado;
- *doNotCheckCapabilities*: Se configurado, as capacidades do classificador não são verificadas antes que o classificador seja compilado (Use com cuidado para reduzir o tempo de execução);
- *epsilon*: O *epsilon* para erro de arredondamento (Não deve ser alterado).

A figura 30 apresenta os resultados obtidos pelo weka na aplicação do modelo SMO.

FIGURA 30: RESULTADOS DO SMO

```
Time taken to build model: 920.71 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      146964           94.4032 %
Incorrectly Classified Instances    8713             5.5968 %
Kappa statistic                    0.8517
Mean absolute error                 0.056
Root mean squared error             0.2366
Relative absolute error             15.8269 %
Root relative squared error         56.2619 %
Total Number of Instances          155677

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,935	0,026	0,992	0,935	0,963	0,858	0,955	0,977	Sim
	0,974	0,065	0,817	0,974	0,889	0,858	0,955	0,802	NÃO
Weighted Avg.	0,944	0,035	0,952	0,944	0,946	0,858	0,955	0,937	

```

=== Confusion Matrix ===

      a      b  <-- classified as
112171  7783 |      a = Sim
  930 34793 |      b = NÃO
```

FONTE: A AUTORA (2019)

Foram classificados corretamente 94,40% das amostras. O Erro Absoluto médio, expressa o erro em uma maneira de fácil compreensão. Medida que indica a média do afastamento de todos os valores fornecidos pelos classificadores e o seu real valor e foi de 0,056; A Raiz quadrada do erro quadrático médio, expressa a extensão da variação dos dados foi de 0,2366, conforme apresentado na figura 31. O resultado encontra-se disposto no Apêndice C.

5.2 COMPARAÇÃO DOS RESULTADOS DOS MODELOS APLICADOS

A tabela 9 demonstra resumidamente a comparação entre os resultados obtidos na base discretizada utilizando os métodos J48, *Multilayer Perceptron* e *SMO*

TABELA 9 - COMPARATIVO DOS RESULTADOS OBTIDOS DOS MODELOS

	J48	MULTILAYER PERCEPTRON	SMO
Instancias	155,667	155,667	155,667
Classificadas Corretas	94.49%	94.47%	94.40%
Incorretas	5.51%	5.53%	5.60%
Tempo de treinamento (s)	1.31	1976.41	920.71
Kappa	0.8548	0.8542	0.8517
Erro Absoluto Médio	0.0898	0.0892	0.056
Erro Quadrado Médio	0.2121	0.2125	0.2366
Erro Absoluto Relativo	25.40%	25.23%	15.83%
Erro relativo ao quadrado da raiz	50.43%	50.53%	56.26%

FONTE: A AUTORA (2019)

O tempo de construção do modelo do J48 é menor com relação aos outros modelos, porém o tempo de treinamento é maior que o modelo SMO e menor que o modelo de RNA. O *Kappa Statistic (Estatística Kappa)* indica o grau de concordância e tem como valor máximo 1, onde 1 representa total concordância e os valores próximos e até abaixo de 0 indicam nenhuma concordância. O J48 obteve o resultado de 0,8548; RNA de 0,8542; e SMO de 0,8517, por mais que o J48 apresente uma estatística maior, todos tiveram resultados acima de 0,5, próximo de 0, indicando concordância. Quanto ao *Mean absolute error (Erro Absoluto Médio)* o algoritmo SMO foi o que apresentou melhor resultados, gerando

menor erro na classificação dos atributos, porém os três apresentaram resultados semelhantes. Quanto ao *Root Mean Squared Error* (Erro Quadrado Médio) o algoritmo SMO apresentou menor erro entre os valores atuais e os valores preditos. Quanto ao Relative Absolute Error (Erro Absoluto Relativo) o algoritmo *Multilayer Perceptron* obteve mais precisão para previsão numérica, ficando bem próximo ao apresentado pelo J48.

Ao analisar a quantidade de classificações erradas, nota-se que J48 apresentou menos instâncias erradas (8.583), contra 8.712 do SMO e 8.610 do RNA.

Comparando a média ponderada (figura 34) obtida nos experimentos é possível identificar que o J48 e *Multilayer Perceptron* apresentaram maior porcentagem de verdadeiros positivos (94,5%) e J48 também apresentou menor porcentagem de falsos positivos (15,0%), o que indica que classificou melhor os atributos. Quanto a precisão, J48 (com 99,50%) mostraram-se um pouco mais efetivo que SMO (com 95,20%) e *Multilayer Perceptron* (com 95,30%). Analisando o F-Measure, SMO e *Multilayer Perceptron* apresentaram resultados iguais (94,60%). Por fim, o ROC Area do J48 alcançou 96,60%, SMO com 95,50% e *Multilayer Perceptron* obteve 96,80%, realizando uma melhor classificação dos atributos. A tabela 10 apresenta a relação média da acurácia dos modelos.

TABELA 10 - MÉDIA DE ACURÁCIA DOS ALGORITMOS

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
J48	93.30%	15.00%	99.50%	93.30%	96.30%	86.20%	96.60%	99.00%
MULTILAYER PERCEPTRON	94.50%	29.00%	95.30%	94.50%	94.60%	86.10%	96.80%	95.40%
SMO	94.40%	35.00%	95.20%	94.40%	94.60%	85.80%	95.50%	93.70%

FONTE: A AUTORA (2019)

Considerando esses aspectos, J48 mostrou-se mais eficiente em quase todos os parâmetros de acurácia, porém *Multilayer Perceptron* resultou em aspectos semelhantes.

6. CONSIDERAÇÕES FINAIS

Com as alterações no comportamento do consumidor devido a inserção de novas tecnologias, a forma de vender e comprar tem sofrido alterações significativas, já que vem migrando cada vez mais para esfera online. Esse comportamento dos consumidores torna o mercado de e-commerce no Brasil muito mais abrangentes e é possível vender para um território muito maior do que antes da era das vendas online, tornando praticável a venda de qualquer coisa em qualquer lugar. Há um aumento significativo no número de empresas que vendem online, com esse aumento da concorrência há um grande aumento na taxa de *churn*, pouco visada pelos e-commerces brasileiros.

Um dos maiores desafios do marketing em e-commerces é a redução da rotatividade dos clientes. O *Churn* é uma métrica comumente utilizada por empresas para avaliar sua retenção de clientes e conseguir projetar com maior confiabilidade seu crescimento nos anos futuros. No Brasil o termo está associado principalmente a empresas que possuem assinatura de um serviço ou produto e vem sendo estudado principalmente na área de telecomunicações. Porém o *churn* não deve ser considerado apenas nesse cenário, já que nesse setor os clientes rejeitam o fornecedor ou compram menos. Como os clientes existentes são uma importante fonte de lucros comerciais, sendo capaz de identificar clientes que mostram sinais de que estão prestes a partir podem criar mais lucros para empresas. Isto é especialmente importante para clientes online, uma vez que o fenômeno do *churn* de clientes parece ser muito rápido e de difícil compreensão. Se as empresas não tomarem medidas para reter os clientes antes que o seu estatuto se deteriore, os clientes podem nunca mais voltar, resultando em desperdício de investimento e perda de ganhos futuros.

Ao traduzir os títulos, palavras-chave plus, palavras-chave e títulos dos documentos os maiores destaques são os tópicos "Previsão", "*Churn*" e "clientes"; e "Dados" e "Padrões", ambos com seis aparições, em relação às Palavras-chave plus dos documentos, ou seja, aquelas acrescentadas pela própria base de dados, os principais tópicos consistem em "Vendas", "Previsão" e "Vendas".

A utilização de tecnologias e de *machine learning* podem ser grandes aliados para essa ação de diminuição da taxa de *churn*. A construção de um mecanismo de

previsão para monitorar o cliente *churn* é um passo importante para o desenvolvimento do negócio, e têm se tornou um tema popular nos últimos dez anos. De um modo geral, os investigadores utilizam a informação sobre comportamentos anteriores de um banco de dados, como demografia de fundo, registros de transações e interações. Toda esta informação é quantificada em variáveis e usado para criar modelos de previsão para prever a probabilidade de perda de clientes no futuro. Metodologias de *machine learning* estão sendo utilizadas de forma preditiva visando a redução do *churn* em empresas.

Dos objetivos específicos, o primeiro: pesquisar na literatura as ferramentas e técnicas que são mais utilizadas para a aplicação de modelo preditivo de *churn*. Foi realizado um levantamento bibliográfico, que constatou que nos últimos dez anos há um aumento no estudo relacionado ao tema, e os países que mais pesquisam sobre são Reino Unido, China e Índia. No Brasil Não há pesquisas que envolvam *machine learning* e e-commerces, talvez devido ao crescimento recente do setor. Constatou-se ao aplicar a lei de Bradford que, mesmo que com baixa quantidade de registros, a distribuição das fontes em zonas corresponde à suposição de que a primeira zona é composta por um pequeno grupo de periódicos mais relacionados ao tema; enquanto que a segunda e a terceira zonas exigem um número maior de periódicos.

As metodologias mais utilizadas envolvem árvore de decisão, *Support Vector Machine* e *Rede Neural Artificial*. *Random Forest*, *Regressão Logística*. A maioria dos estudos existentes relacionados com a previsão do *churn* de clientes são análises estáticas e não estão bem adequados à realização de monitorizações individuais e dinâmicas. A análise de dados estáticas não fornecem monitoramento dinâmico do *churn* do cliente. Dada a natureza extremamente volátil dos mercados em linha, seria desejável adotar uma abordagem mais dinâmica. Além disso, os resultados de análise de dados estáticos repetidos são exibidos em tabelas massivas, e engenheiros precisam fazer esforços consideráveis para analisar as diferenças preditivas.

Outro objetivo era definir métodos de *machine learning* para predição de desistência de compra de cliente. Após a análise bibliográfica notou-se uma metodologia comum aplicada em praticamente todas as pesquisas. Primeiro realizavam o pré-processamento da base de dados, retirados normalmente de um *datawarehouse* (DW), depois seguiam para a classificação dos clientes,

clusterizando-os normalmente em dois grupos (VIPs e Não VIPs), e por último buscavam regras de associação e padrões sequenciais aplicando alguns modelos de *machine learning* para predição.

O último objetivo consiste em analisar as principais ferramentas e técnicas visando sua aplicabilidade no modelo de negócio de um e-commerce. Foi utilizado uma base de dados real para teste dos algoritmos de *machine learning* mais utilizados, conforme selecionado após o levantamento bibliográfico. Os métodos escolhidos foram: Árvore de Decisão - J48, Rede Neural Artificial - *Multilayer Perceptron* e *Support Vector Machine* - SMO. A base trata de um e-commerce na área de varejo, um dos maiores marketplaces de produtos para casa do Brasil, que conta com diversas categorias e mais de um milhão de produtos. Os três modelos apresentaram resultados similares quando analisados a taxa de acertos do modelo, porém o que apresenta melhor tempo de execução do modelo (1,31 segundos) é a Árvore de Decisão J48.

Há poucas pesquisas relacionadas com o tema no mercado brasileiro, e poucos trabalhos que relacionam técnicas de *machine learning* com predição dentro de e-commerces. Esta pesquisa também compreende a possibilidade de pesquisas que aprofundem os impactos e os aspectos de uns fatores encontrados. Recomenda-se a aplicação de outros modelos preditivos além dos apresentados, visando encontrar o melhor modelo para os dados a serem treinados.

REFERÊNCIAS

AGRAWAL, Sanket; DAS, Aditya; GAIKWAD, Amit; DHAGE, Sudhir. **Customer churn prediction modelling based on behavioural patterns analysis using deep learning**. In: International Conference on Smart computing and electronic enterprise, Malásia, Julho de 2018. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8538420>>. Acesso em: 01 jun. 2019

ALBUQUERQUE, P., ALFINITO, S., TORRES, C.V.: Support vector clustering for customer segmentation on mobile tv service. **Communications in Statistics – Simulation and Computation**, v.44 p.212–216. 2015. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/03610918.2013.794289>>. Acesso em: 17 nov. 2019.

ARIA, M.; CUCCURULLO, C. **Bibliometrix**: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11(4), pp 959-975, Elsevier, 2017.

BAEKDAL, Thomas. **Where is Everyone?** 2009. Disponível em: <<https://www.baekdal.com/strategy/market-of-information>>. Acesso em: 08 dez. 2019

BOENTE, A. N. P. ; GOLDSCHMIDT ; ESTRELA, V. V.. Uma Metodologia de Suporte ao Processo de Descoberta de Conhecimento em Bases de Dados. In: **V Simpósio de Excelência em Gestão e Tecnologia**, 2008, Resende - RJ. V SEGeT, 2008. v. 1. p. 4-5. Disponível em: <<http://boente.eti.br/publica/seget2008kdd.pdf>> Acesso em: 17 nov. 2019

BRAGA, A. de.; LUDERMIR, T. B.; CARVALHO, A. C. P. de L. F. **Redes Neurais Artificiais teoria e aplicações**, Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A., 2000

Caldeira, G. Brandao, A. C. M. Pereira, "Fraud Analysis and Prevention in e-Commerce Transactions", **9th Latin American Web Congress**, pp. 42-49, 2014.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Goiânia: Ufg, 2009.

CARDOSO, D.L.; MANGANOTE, E.J.T.: **Empresas virtuais**: Tópicos atuais de administração, São Paulo: Alínea, 2000.

CASTRO, L. N.; FERRARI, D. G.. **Introdução à mineração de dados**: conceitos básicos, algoritmos e aplicações. 1a Edição. São Paulo: Saraiva, 2016

CHAMBERLAIN, Benjamin Paul; et. al.. Customer lifetime value prediction using embeddings. **International conference on knowledge discovery and data Mining** p. 1753-1762, 2017. Disponível em: < <https://arxiv.org/pdf/1703.02596.pdf>>. Acesso em: 01 jun. 2019

CHEN, Ssu-Han. The gamma CUSUM chart method for online customer *churn* prediction. **Journal Elsevier - Expert Systems with Applications**, Amsterdã, v. 39, n. 9, p. 99-111, 2012. Disponível em: <<https://daneshyari.com/article/preview/379552.pdf>>. Acesso em: 01 jun. 2019

CHEN, Zhen-Yu; FAN, Zhi-Ping; SUN, Minghe. **A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data**. In: European Journal of Operational Research, Elsevier, v. 223, n.2, p. 461-472, 2012. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S037722171200495X>>. Acesso em: 01 jun. 2019

CHEESEMAN, P.; J. STUTZ. Bayesian classification (Autoclass): Theory and results advances in knowledge discovery and data mining. **NASA**. Disponível em: <https://www.researchgate.net/publication/2659256_Bayesian_ClassificationAutoClassTheory_and_Results> Acesso em: 17 nov. 2019

E-BIT, Notícias da. **WEBSHOPPERS**. 39 eds. 2019 Disponível em: <<https://www.ebit.com.br/webshoppers>>. Acesso em: 12 jun. 2019

FAYYAD, U. M.; et. Al.. From Data Mining to Knowledge Discovery in Databases. **Artificial Intelligence Magazine**, v. 17, n. 3, p. 37-54, 1996.

FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002. Apostila. Disponível em: <<http://www.ia.ufrj.br/ppgea/conteudo/conteudo-2012-1/1SF/Sandra/apostilaMetodologia.pdf>> Acesso em: 14 jun. 2019.

GABRIEL, M. **Marketing na era digital**: conceitos, plataformas e estratégias. São Paulo: Novatec, 2010.

GARCIA-RAMON, M. D.; MONK, J. **Gender and geography**: World views and practices. *Belgeo*, n. 3, p. 247–259, 2007.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

GLADY, N; et al.; Modeling *churn* using customer lifetime value. **European Journal of Operational Research**, v. 197, n. 1, p. 402–411, 2009

HADDEN, John et al. Computer assisted customer churn management: State-of-the-art and future trends. **Computers & Operations Research**, v. 34, n. 10, p.2902-2917, 2007. Disponível em: <https://www.researchgate.net/publication/222701625_Computer_assisted_customer_churn_management_State-of-the-art_and_future_trends>. Acesso em: 17 nov. 2019.

Haykin, S. (1999). **Neural Networks: a Comprehensive Foundation**. Prentice Hall. Disponível em: <https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural%20Networks%20-%20A%20Comprehensive%20Foundation%20-%20Simon%20Haykin.pdf> Acesso em: 08 Dez. 2019

Holsheimer, M; Siebes, A. P. Data Mining: The Search for Knowledge in Databases. **Centre for Mathematics and Computer Science**. Box 94079, Amsterdam, 1994 Disponível em: <<https://pdfs.semanticscholar.org/3889/071d26158b7a1c52620e517ffd6f4df32789.pdf>> Acesso em: 14 nov. 2019

KERAMATI, Abbas; et. al. Developing a prediction model for customer *churn* from electronic banking services using data mining. **Financial Innovation**, v. 2, n. 1, p. 1-13, 2016. Disponível em: <https://link.springer.com/content/pdf/10.1186%2Fs40854-016-0029-6.pdf>. Acesso em: 01 jun. 2019

KIM, Kyoungok; LEE, Jaewook. Sequential manifold learning for efficient *churn* prediction. **Journal Elsevier - Expert Systems with Applications**, Amsterdã, v.39, n.9, p. 13328–13337, 2012. Disponível em: https://www.researchgate.net/publication/257404518_Sequential_manifold_learning_for_efficient_churn_prediction. Acesso em: 01 jun. 2019

KOTLER, P. **Administração de Marketing**. 10. ed. São Paulo: Person, 2000.

KOTLER, P., KARTAJAYA, H., SETIWAN, I.. **Marketing 4.0**: do tradicional ao digital. Rio de Janeiro: Sextante, 2017.

KOTSIANTIS, S. B. "Supervised machine learning: a review of classification techniques, **Informatica** 31, 249–268, 2007. Disponível em: <[https://datajobs.com/data-science-repo/Supervised-Learning-\[SB-Kotsiantis\].pdf](https://datajobs.com/data-science-repo/Supervised-Learning-[SB-Kotsiantis].pdf)> Acesso em: 17 nov 2019

KUMAR V., REINARTZ W. Impact of CRM on Marketing Channels. In: Customer Relationship Management. Springer Texts in Business and Economics. Springer, Berlin, Heidelberg, 2012

LAZAROV, V., CAPOTA, M. **Churn prediction**, Business Analytics Course. Munique: TUM Computer Science. 2007. Disponível em:

<<https://pdfs.semanticscholar.org/dbf1/5b7c5f766ef9f84ba83127c626d79b2087b2.pdf>> Acesso em: 12 jun. 2019

LEJEUNE, M. (2001). Measuring the impact of data mining on *churn* management. **Internet Research**, 2001 11(6), 375 -387. Disponível em: <<https://pdfs.semanticscholar.org/262f/dce2491db4b98e70ece38055d6ef31774c31.pdf>> Acesso em: 12 jun. 2019

MALHOTRA, Naresh K.. **Pesquisa de marketing**: uma orientação aplicada. 4. ed. Porto Alegre: Bookman, 2006. 720 p.

MITCHELL, T. M. **machine learning**. New York: McGraw-Hill, 1997. Disponível em: <<http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf>>. Acesso em: 03 nov. 2019.

MOFAKHAN, Faria; et. Al. Electronic promotion to new customers using mk NN learning. **Information sciences**, Sharjah, v. 179 , p. 248-266, 2008. Disponível em: <https://www.academia.edu/304437/Electronic_Promotion_to_New_Customers_using_mkNN_Learning?auto=download>. Acesso em: 01 jun. 2019

MOHER, D.; et Al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **Annals of internal medicine**, v. 151, n. 4, p. 264–269, 2009. Disponível em <<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097>> Acesso em: 01 nov. 2019

MONARD, M.C.; BARANAUSKAS, J.A. Conceitos sobre aprendizado de máquina. In: Rezende, S.O. (Ed.). **Sistemas Inteligentes**: Fundamentos e aplicações. São Carlos, 2003. p.89-114. cap. 4.

NESLIN, S. A.; et. Al. Defection detection: measuring and understanding the predictive accuracy of customer *churn* models. **Journal of Marketing Research**, Chicago, Vol. 43, n.2, 2006. Disponível em: <https://www.researchgate.net/publication/240474361_Challenges_and_Opportunities_in_Multichannel_Customer_Management/citation/download> Acesso em: 17 nov. 2019

O'BRIEN, James A. **Sistemas de Informação**: e as decisões gerenciais na era da Internet. 2. ed. São Paulo: Saraiva, 2004.

PARAISO, Gustavo José Barbosa. **O E-commerce nas redes sociais**: estudo sobre os desdobramentos do comércio eletrônico na atualidade. Pernambuco: UPE, 2011. Disponível em:

<<http://www.bocc.uff.br/pag/paraiso-gustavo-o-e-commerce-nas-redes-sociais.pdf>>. Acesso em: 08 jun. 2019.

PEREIRA, Jorge Luis. **ANÁLISE PREDITIVA EM SISTEMAS DE INFORMAÇÃO NO CONTEXTO DO BIG DATA**. 2014. 72 f. Tese (Doutorado) - Curso de Bacharelado em Sistemas de Informação, Centro Universitário Eurípides de Marília, Marília, 2015. Disponível em: <<https://aberto.univem.edu.br/bitstream/handle/11077/1003/Jorge%20Lu%c3%ads%20Pereira.pdf?sequence=1&isAllowed=y>>. Acesso em: 08 dez. 2019.

RACHID, Ait Daoud; ABDELLAH, Amine; BELAID, Bouikhalene; RACHID, Lbibb. Clustering Prediction Techniques in Defining and Predicting Customers Defection: The Case of E-Commerce Context. **International Journal of Electrical and Computer Engineering**, Marrocos, v. 8, n. 4, p. 2367-2383, 2018. Disponível em: <https://www.researchgate.net/profile/Rachid_Ait_Daoud/publication/324149924_Clustering_Prediction_Techniques_in_Defining_and_Predicting_Customers_Defection_The_Case_of_E-Commerce_Context/links/5b433766a6fdccbcf90e0e13/Clustering-Prediction-Techniques-in-Defining-and-Predicting-Customers-Defection-The-Case-of-E-Commerce-Context.pdf>. Acesso em: 01 jun. 2019

RAUBER, T. W. Redes Neurais Artificiais. **ERI - Encontro Regional de Informática**. Nova Friburgo: Sociedade Brasileira de Computação, p. 201-228, 1998. Disponível em: <https://www.researchgate.net/profile/Thomas_Rauber/publication/228686464_Redes_neurais_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf> Acesso em: 15 nov. 2019

ROJAS, R. **Neural Networks**: A systematic Introduction. Springer-Verlag, Berlin, 1996. Disponível em: <<https://page.mi.fu-berlin.de/rojas/neural/neuron.pdf>> Acesso em: 28 out. 2019

RUIZ, R.S.R et al. Árvores de Decisão na Classificação de Dados Astronômicos. **TEMA - Tendências em Matemática Aplicada e Computacional**, [S.l.], v. 10, n. 1, p. 75-86, june 2009. ISSN 2179-8451. Disponível em: <<https://tema.sbmac.org.br/tema/article/view/126>>. Acesso em: 13 nov. 2019.

SHIM, Beomsoo; et Al.. CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns. **Journal Elsevier - Expert Systems with Applications**, Amsterdã, v.39, n.9, p. 7736-7742, 2012. Disponível em: <https://www.matlabi.ir/wp-content/uploads/bank_papers/ipaper/o41_www.Matlabi.ir_CRM%20strategies%20for%20a%20small-sized%20online%20shopping%20mall%20based%20on%20association%20rules%20and%20sequential%20patterns.pdf> Acesso em: 01 jun. 2019

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis, 2005.

SUBRAMANYA, Karthik; SOMANI, Arun. **Enhanced Feature Mining and Classifier Models to Predict Customer Churn for an E-retailer**. In: International Conference on Cloud Computing, Data Science & Engineering – Confluence, Índia, p. 531-536, 2017. Disponível em: <https://ieeexplore.ieee.org/document/7943208>. Acesso em: 01 jun. 2019

TURCHI, Sandra R. **Estratégias de marketing digital e e-commerce**. São Paulo: Editora Atlas, 2012.

VALERO-FERNANDEZ, et. Al.. Towards Accurate Predictions of Customer Purchasing Patterns. **International Conference on Computer and Information Technology**, Finlândia, p. 157-161, 2017. Disponível em: <https://ieeexplore.ieee.org/document/8031468>. Acesso em: 01 jun. 2019

VAZ, Adolpho Conrado. **Os 8Ps do marketing digital: o seu guia estratégico de marketing digital**. São Paulo: Novatec Editora, 2011.

WANG, Qiu-Feng; XU, Mirror; HASSAIN, Amir. Large-scale Ensemble Model for Customer Churn Prediction in Search Ads. **Cognitive Computation**, Suíça, v. 11, n.2, 2018. Disponível em: <https://link.springer.com/article/10.1007/s12559-018-9608-3>. Acesso em: 01 jun. 2019

WU, Xiaojun; MENG, Sufang. **E-commerce Customer Churn Prediction Based on Improved SMOTE and AdaBoost**. In: International Conference on Service Systems and Service Management, China, p. 1-5, 2016. Disponível em: <https://ieeexplore.ieee.org/document/7538581>. Acesso em: 01 jun. 2019

YU, Xiaobing; et Al.. An extended *support vector machine forecasting framework* for customer churn in e-commerce. **Expert Systems with Applications**, China, V. 38, n.3, p. 1425–1430, 2011. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417410006779?via%3Dihub>. Acesso em: 02 jun. 2019

APÊNDICE A – LISTA DE ARTIGOS

TÍTULO	AUTORES	LINK
CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns	Shim B;Choi K;Suh Y	https://www.matlabi.ir/wp-content/uploads/bank_papers/ipaper/o41_www.Matlabi.ir_CRM%20strategies%20for%20a%20small-sized%20online%20shopping%20mall%20based%20on%20association%20rules%20and%20sequential%20patterns.pdf
Clustering prediction techniques in defining and predicting customers defection: the case of e-commerce context	Rachid Ad;Abdellah A;Belaid B;Rachid L	https://www.researchgate.net/profile/Rachid_Ait_Daoud/publication/324149924_Clustering_Prediction_Techniques_in_Defining_and_Predicting_Customers_Defection_The_Case_of_E-Commerce_Context/links/5b433766a6fdccbcf90e0e13/Clustering-Prediction-Techniques-in-Defining-and-Predicting-Customers-Defection-The-Case-of-E-Commerce-Context.pdf
Customer <i>churn</i> prediction modelling based on behavioural patterns analysis using deep learning	Agrawal S;Das A;Gaikwad A;Dhage S	https://ieeexplore.ieee.org/abstract/document/8538420
Electronic promotion to new customers using mknn learning	Nassiri Mofakham F;Nematbakhsh Ma;Baraani Dastjerdi A;Ghasem Aghaee N	https://www.academia.edu/304437/Electronic_Promotion_to_New_Customers_using_mkNN_Learning?auto=download
Customer lifetime value prediction using embeddings	Chamberlain Bp;Cardoso;Bryan Liu Ch;Pagliari R;Deisenroth Mp	https://arxiv.org/pdf/1703.02596.pdf
Sequential manifold learning for efficient <i>churn</i> prediction	Kim K;Lee J	https://www.researchgate.net/publication/257404518_Sequential_manifold_learning_for_efficient_churn_prediction
The gamma cusum chart method for online customer <i>churn</i>	Chen Sh	https://daneshyari.com/article/preview/379552.pdf

prediction		
Developing a prediction model for customer <i>churn</i> from electronic banking services using data mining	Keramati A;Ghaneei H;Mirmohammadi Sm	https://link.springer.com/content/pdf/10.1186%2Fs40854-016-0029-6.pdf
A hierarchical multiple kernel <i>support vector machine</i> for customer <i>churn</i> prediction using longitudinal behavioral data	Chen Zy;Fan Zp;Sun M	https://www.sciencedirect.com/science/article/abs/pii/S037722171200495X
Data mining techniques in customer <i>churn</i> prediction	TSAI CF;LU YH	https://pdfs.semanticscholar.org/9ee8/0a16e173d7604318cd80cd5509abfda3e932.pdf?_ga=2.178710180.148694644.1558696740-1699176466.1558696740
Enhanced feature mining and classifier models to predict customer <i>churn</i> for an e-retailer	Subramanya Kb;Somani A	https://ieeexplore.ieee.org/document/7943208
Combining local and social network classifiers to improve <i>churn</i> prediction	Backiel A;Verbinnen Y;Baesens B;Claeskens G	https://ieeexplore.ieee.org/document/7403612
Towards accurate predictions of customer purchasing patterns	Valero Fernandez R;Collins Dj;Lam Kp;Rigby C;Bailey J	https://ieeexplore.ieee.org/document/8031468
Large-scale ensemble model for customer <i>churn</i> prediction in search ads	Wang Qf;Xu M;Hussain A	https://link.springer.com/article/10.1007/s12559-018-9608-3
E-commerce customer <i>churn</i> prediction based on improved smote and adaboost	Wu X;Meng S	https://ieeexplore.ieee.org/document/7538581
Research on e-commerce user <i>churn</i> prediction based on logistic regression	Yanfang Q;Chen L	https://ieeexplore.ieee.org/document/8284914
An extended <i>support vector</i>	Yu X;Guo S;Guo	https://www.sciencedirect.com/science/article/p

<i>machine forecasting framework</i> for customer <i>churn</i> in e-commerce	J;Huang X	ii/S0957417410006779?via%3Dihub
Challenges in the analysis of online social networks: a data collection tool perspective	Goswami A;Kumar A	https://link.springer.com/article/10.1007/s11277-017-4712-3
Generic framework to predict repeat behavior of customers using their transaction history	Kazmi Ah;Shroff G;Agarwal P	https://ieeexplore.ieee.org/document/7817089
Predicting behavior	Abbasi A;Lau Ryk;Brown De	https://scholars.cityu.edu.hk/en/publications/publication(1c1d2cf5-5b0e-4b99-9194-b48a2a5b1129).html
An empirical comparison of techniques for the class imbalance problem in <i>churn</i> prediction	Zhu B;Baesens B;Vanden Broucke Sklm	https://www.researchgate.net/publication/316028407_An_empirical_comparison_of_techniques_for_the_class_imbalance_problem_in_churn_prediction

APÊNDICE B - RESULTADO DO ALGORITMO J48

```
==== Run information ====
```

```
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
```

```
Relation: base_cliente54
```

```
Instances: 155677
```

```
Attributes: 9
```

```
tipopessoa
```

```
estado
```

```
newsletter
```

```
valorgasto
```

```
pedidos
```

```
taxaaprovacao
```

```
cadastroprimeiracompra
```

```
primeiraultimacompra
```

```
churn
```

```
Test mode: 10-fold cross-validation
```

```
==== Classifier model (full training set) ====
```

```
J48 pruned tree
```

```
-----
```

```
newsletter <= 0
```

```
| primeiraultimacompra = A
```

```
| | cadastroprimeiracompra = A
```

```
| | | tipopessoa = F
```

```
| | | | taxaaprovacao = C: Sim (19029.0/1.0)
```

```
| | | | taxaaprovacao = B: Não (3.0)
```

```
| | | | taxaaprovacao = A: Sim (86303.0/33.0)
```

```
| | | tipopessoa = J: Sim (1104.0)
```

```
| | | tipopessoa = I: Não (3.0)
```

```
| | cadastroprimeiracompra = B: Não (15.0/4.0)
```

```
| | cadastroprimeiracompra = C: Não (17.0/6.0)
```

```
| primeiraultimacompra = C: Não (39.0/8.0)
```

```
| primeiraultimacompra = B
```

```
| | tipopessoa = F: Não (41.0/11.0)
```

```

| | tipopessoa = J: Sim (2.0)
| | tipopessoa = I: NÃo (1.0)
newsletter > 0
| taxaaprovacao = C
| | primeiraultimacompra = A
| | | valorgasto = C: Sim (4.0/1.0)
| | | valorgasto = B: NÃo (7.0/2.0)
| | | valorgasto = A: Sim (5830.0/444.0)
| | primeiraultimacompra = C: NÃo (267.0/105.0)
| | primeiraultimacompra = B
| | | tipopessoa = F: NÃo (314.0/139.0)
| | | tipopessoa = J: NÃo (2.0)
| | | tipopessoa = I: Sim (3.0)
| taxaaprovacao = B
| | tipopessoa = F: NÃo (2570.0/332.0)
| | tipopessoa = J: NÃo (44.0/9.0)
| | tipopessoa = I: Sim (13.0/4.0)
| taxaaprovacao = A
| | tipopessoa = F
| | | valorgasto = C: NÃo (0.0)
| | | valorgasto = B: Sim (4.0/1.0)
| | | valorgasto = A: NÃo (39591.0/7355.0)
| | tipopessoa = J: NÃo (340.0/76.0)
| | tipopessoa = I: Sim (131.0/32.0)

```

Number of Leaves : 26

Size of the tree : 39

Time taken to build model: 1.31 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	147093	94.486 %
Incorrectly Classified Instances	8584	5.514 %
Kappa statistic	0.8548	
Mean absolute error	0.0898	

Root mean squared error	0.2121
Relative absolute error	25.4005 %
Root relative squared error	50.4305 %
Total Number of Instances	155677

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,933	0,015	0,995	0,933	0,963	0,862	0,966	0,990	Sim
	0,985	0,067	0,814	0,985	0,891	0,862	0,966	0,820	NÃo
Weighted Avg.	0,945	0,027	0,954	0,945	0,947	0,862	0,966	0,951	

=== Confusion Matrix ===

```

a    b  <-- classified as
111896 8058 |    a = Sim
 526 35197 |    b = NÃo

```

APÊNDICE C - RESULTADO DO SMO

=== Run information ===

Scheme: weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K

"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator

"weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"

Relation: base_cliente54

Instances: 155677

Attributes: 9

tipopessoa

estado

newsletter

valorgasto

pedidos

taxaaprovacao

cadastroprimeiracompra

primeiraultimacompra

churn

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

SMO

Kernel used:

Linear Kernel: $K(x,y) = \langle x,y \rangle$

Classifier for classes: Sim, Não

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.6666 * (normalized) tipopessoa=F
 + 0.6667 * (normalized) tipopessoa=J
 + -1.3333 * (normalized) tipopessoa=I
 + 0.0383 * (normalized) estado=PR
 + 0.0389 * (normalized) estado=SP
 + 0.0383 * (normalized) estado=MG
 + 0.0386 * (normalized) estado=ES
 + 0.0385 * (normalized) estado=SC
 + 0.0384 * (normalized) estado=RS
 + 0.0385 * (normalized) estado=RJ
 + 0.038 * (normalized) estado=PI
 + 0.0385 * (normalized) estado=GO
 + 0.038 * (normalized) estado=CE
 + 0.0387 * (normalized) estado=SE
 + 0.0386 * (normalized) estado=DF
 + 0.038 * (normalized) estado=PE
 + 0.0387 * (normalized) estado=BA
 + 0.0383 * (normalized) estado=AM
 + 0.0381 * (normalized) estado=AL
 + 0.0381 * (normalized) estado=MT
 + 0.0389 * (normalized) estado=MS
 + 0.0381 * (normalized) estado=PB
 + 0.0379 * (normalized) estado=TO
 + 0.0389 * (normalized) estado=MA
 + 0.038 * (normalized) estado=PA

```

+ 0.0391 * (normalized) estado=RN
+ 0.0387 * (normalized) estado=RO
+ -1 * (normalized) estado=AC
+ 0.039 * (normalized) estado=RR
+ 0.0388 * (normalized) estado=AP
+ 2.0003 * (normalized) newsletter
+ 0.0005 * (normalized) valorgasto=B
+ -0.0005 * (normalized) valorgasto=A
+ -1.3333 * (normalized) pedidos=A
+ 0.6661 * (normalized) pedidos=B
+ 0.6672 * (normalized) pedidos=C
+ -1.3335 * (normalized) taxaaprovacao=C
+ 0.6667 * (normalized) taxaaprovacao=B
+ 0.6668 * (normalized) taxaaprovacao=A
+ 0.0001 * (normalized) cadastroprimeiracompra=A
+ -0.0001 * (normalized) cadastroprimeiracompra=B
+ -0 * (normalized) cadastroprimeiracompra=C
+ 0.0002 * (normalized) primeiraultimacompra=A
+ -0.0002 * (normalized) primeiraultimacompra=C
+ 0.0001 * (normalized) primeiraultimacompra=B
- 1.0384

```

Number of kernel evaluations: -1902971448

Time taken to build model: 920.71 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	146964	94.4032 %
Incorrectly Classified Instances	8713	5.5968 %
Kappa statistic	0.8517	
Mean absolute error	0.056	
Root mean squared error	0.2366	
Relative absolute error	15.8269 %	
Root relative squared error	56.2619 %	
Total Number of Instances	155677	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,935	0,026	0,992	0,935	0,963	0,858	0,955	0,977	Sim
	0,974	0,065	0,817	0,974	0,889	0,858	0,955	0,802	Não
Weighted Avg.	0,944	0,035	0,952	0,944	0,946	0,858	0,955	0,937	

=== Confusion Matrix ===

```

a    b  <-- classified as
112171 7783 |  a = Sim
 930 34793 |  b = Não

```

APÊNDICE D – RESULTADO DO MULTILAYER PERCEPTRON

=== Run information ===

Scheme: weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Relation: base_cliente54

Instances: 155677

Attributes: 9

tipopessoa

estado

newsletter

valorgasto

pedidos

taxaaprovacao

cadastroprimeiracompra

primeiraultimacompra

churn

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Sigmoid Node 0

Inputs Weights

Threshold -1.3436577231078424

Node 2 0.5475915829800894

Node 3 -5.199425190186672

Node 4 0.5263126579735752

Node 5 0.6613523683369404

Node 6 0.3968120240975882

Node 7 0.9965134880024483

Node 8 4.339387373467507

Node 9 -5.780204278876824

Node 10 2.43598144766964

Node 11 0.696385597188814

Node 12 0.12039594673867539

Node 13 0.9108008455944195

Node 14 9.504669259543974

Node 15 1.8169289975314789

Node 16 3.248105888453486

Node 17 2.9779181457060164

Node 18 0.3575967032591999

Node 19 2.8571713110563044

Node 20 2.8838607355976253

Node 21 0.9722725513491536

Node 22 1.1547664965465727

Node 23 5.142882505882065

Node 24 -0.8366097953069302

Node 25 3.101922539371896

Node 26 2.9035449175488983

Node 27 0.853541136107727

Sigmoid Node 1

Inputs Weights

Threshold 1.3436577039038429

Node 2 -0.5475938772662795

Node 3 5.199359509946107

Node 4 -0.526312978617199

Node 5 -0.6613478849634769

Node 6 -0.39681405977895656

Node 7 -0.9965229844917963

Node 8 -4.335493281405447

Node 9 5.779322161275088
 Node 10 -2.4355440019199723
 Node 11 -0.6964311299617181
 Node 12 -0.12039592392715437
 Node 13 -0.9107901499428791
 Node 14 -9.5041925612429
 Node 15 -1.816898664360535
 Node 16 -3.24874034219614
 Node 17 -2.981706471859911
 Node 18 -0.3575961383518019
 Node 19 -2.8571288163400386
 Node 20 -2.883862360616723
 Node 21 -0.9722563697895552
 Node 22 -1.1547677810623136
 Node 23 -5.143625062787901
 Node 24 0.8366098992910348
 Node 25 -3.101911258819879
 Node 26 -2.8697997011548875
 Node 27 -0.8535391242212197

Sigmoid Node 2

Inputs Weights

Threshold -0.9468391228797888
 Attrib tipopessoa=F -1.3504726514745267
 Attrib tipopessoa=J 2.286851296250118
 Attrib tipopessoa=I 0.015806951712005933
 Attrib estado=PR 3.601480967283757
 Attrib estado=SP 0.5658871946041685
 Attrib estado=MG -1.8793397507849434
 Attrib estado=ES 1.0887035156910292
 Attrib estado=SC -0.16481169100272722
 Attrib estado=RS 5.138920365511989
 Attrib estado=RJ -2.240360681706604
 Attrib estado=PI 1.3009689629088927
 Attrib estado=GO 0.6051569427574078
 Attrib estado=CE 3.390131116236821
 Attrib estado=SE -0.4160203558690298
 Attrib estado=DF -0.5412769913904413
 Attrib estado=PE -0.9466130764140965
 Attrib estado=BA -1.3646197730207015
 Attrib estado=AM 0.28864070506642137
 Attrib estado=AL -0.16215143358428125
 Attrib estado=MT -0.6452539977921463
 Attrib estado=MS -0.6882807166381311
 Attrib estado=PB 0.6650703613089436
 Attrib estado=TO -0.3114694078321402
 Attrib estado=MA 0.30013343371487033
 Attrib estado=PA 0.23309925313930813
 Attrib estado=RN 0.09324657189121831
 Attrib estado=RO -0.3958676273439127
 Attrib estado=AC 0.38502402858752305
 Attrib estado=RR 0.31460766155868375
 Attrib estado=AP 0.365877733076552
 Attrib estado=Sp 0.30018818462977903
 Attrib estado=sp 0.29485000313008636
 Attrib estado=mg 0.3358929454253853
 Attrib estado=pr 0.2908161454463444
 Attrib estado=Es 0.3581377276077341
 Attrib newsletter 0.34477496366725424
 Attrib valorgasto=C 0.9244757456239183
 Attrib valorgasto=B 0.8364535996243305

Attrib valorgasto=A -0.7792915096536164
 Attrib pedidos=A -0.6460939216328219
 Attrib pedidos=B 0.6851399433506289
 Attrib pedidos=C 0.8581316538043866
 Attrib taxaaprovacao=C -1.659793885125919
 Attrib taxaaprovacao=B 1.5987762164043378
 Attrib taxaaprovacao=A 1.05058721617581
 Attrib cadastroprimeiracompra=A -2.758109898570036
 Attrib cadastroprimeiracompra=B -1.1058559135928276
 Attrib cadastroprimeiracompra=C 4.726799230684682
 Attrib primeiraultimacompra=A 2.8100811979997715
 Attrib primeiraultimacompra=C 0.552434321518941
 Attrib primeiraultimacompra=B -2.452871912777396

Sigmoid Node 3

Inputs Weights

Threshold -0.607035179852893
 Attrib tipopessoa=F 1.078439747292299
 Attrib tipopessoa=J -0.2675038610912661
 Attrib tipopessoa=I -0.1795564261310572
 Attrib estado=PR 0.2627150126903599
 Attrib estado=SP -3.2675761810938404
 Attrib estado=MG 0.446713705614829
 Attrib estado=ES -0.4299099163955339
 Attrib estado=SC 1.0729701728384737
 Attrib estado=RS 4.278149387434652
 Attrib estado=RJ -1.4504950303438824
 Attrib estado=PI 0.3488489177036726
 Attrib estado=GO -0.5568188448703189
 Attrib estado=CE 3.2576687983193406
 Attrib estado=SE -0.31644975691003857
 Attrib estado=DF -1.179049153090584
 Attrib estado=PE -0.8411879435627347
 Attrib estado=BA 0.22531179505122656
 Attrib estado=AM 0.1298510985203188
 Attrib estado=AL 0.11204795709238302
 Attrib estado=MT 1.3583510328359123
 Attrib estado=MS -0.5237343641827381
 Attrib estado=PB 0.42239726947730083
 Attrib estado=TO 0.5372848914705018
 Attrib estado=MA 0.04960465977438118
 Attrib estado=PA -0.05147638990696963
 Attrib estado=RN -0.18479092324353108
 Attrib estado=RO 0.7810828561436122
 Attrib estado=AC 0.46043983621101037
 Attrib estado=RR 0.15518427080883598
 Attrib estado=AP 0.1746522105623613
 Attrib estado=Sp 0.2481751893500416
 Attrib estado=sp 0.16776715722744315
 Attrib estado=mg 0.18187070313735343
 Attrib estado=pr 0.16311563019956865
 Attrib estado=Es 0.2247224603818127
 Attrib newsletter 1.0264814565133713
 Attrib valorgasto=C 1.5593303398124028
 Attrib valorgasto=B 2.8579574374473182
 Attrib valorgasto=A -3.689596092612913
 Attrib pedidos=A -2.779405694431203
 Attrib pedidos=B 0.10795144400836501
 Attrib pedidos=C 3.346480666072384
 Attrib taxaaprovacao=C 1.302792278968453
 Attrib taxaaprovacao=B 0.6810502488729706

Attrib taxaaprovacao=A -1.3331664186072618
 Attrib cadastroprimeiracompra=A 1.1259211116957828
 Attrib cadastroprimeiracompra=B -0.1980009386286598
 Attrib cadastroprimeiracompra=C -0.32800074201292684
 Attrib primeiraultimacompra=A 0.7643882962610444
 Attrib primeiraultimacompra=C 1.093605692998824
 Attrib primeiraultimacompra=B -1.2557979638791692

Sigmoid Node 4

Inputs Weights
 Threshold -0.8434038962486774
 Attrib tipopessoa=F -0.2616298538453612
 Attrib tipopessoa=J 4.306203764446966
 Attrib tipopessoa=I -3.1411134739492033
 Attrib estado=PR -5.513707439295155
 Attrib estado=SP -0.7249696271921067
 Attrib estado=MG -0.43491705832164884
 Attrib estado=ES -5.112899951175362
 Attrib estado=SC -5.799750095901838
 Attrib estado=RS 7.483846043424353
 Attrib estado=RJ 2.513938125388657
 Attrib estado=PI 5.700213436540113
 Attrib estado=GO -0.5142137995040603
 Attrib estado=CE -4.472540854483803
 Attrib estado=SE -0.5139547244841711
 Attrib estado=DF -1.3882657544705161
 Attrib estado=PE -4.850897326056016
 Attrib estado=BA 2.0825074992980364
 Attrib estado=AM 2.203467457551719
 Attrib estado=AL -3.866987930297515
 Attrib estado=MT 0.041678923152951665
 Attrib estado=MS 2.7501315505989097
 Attrib estado=PB 3.5951107408217493
 Attrib estado=TO 7.929690615001104
 Attrib estado=MA 2.7940520818278305
 Attrib estado=PA 0.6745276589689938
 Attrib estado=RN -0.38240366790453845
 Attrib estado=RO -0.5509178877077077
 Attrib estado=AC 0.3631049619605823
 Attrib estado=RR 0.2860888792901582
 Attrib estado=AP 0.4817920643805753
 Attrib estado=Sp 0.19414516312793503
 Attrib estado=sp 0.19046971054678144
 Attrib estado=mg 0.22333059638169608
 Attrib estado=pr 0.1529680186682971
 Attrib estado=Es 0.1712014920736257
 Attrib newsletter -7.445838712781343
 Attrib valorgasto=C -0.3769015185648007
 Attrib valorgasto=B 1.3182255575236306
 Attrib valorgasto=A -0.03314367348289817
 Attrib pedidos=A 0.8251823946526077
 Attrib pedidos=B -0.570436132359399
 Attrib pedidos=C 0.5439051003911233
 Attrib taxaaprovacao=C 3.3198998815647562
 Attrib taxaaprovacao=B 1.1327531445288825
 Attrib taxaaprovacao=A -3.5854880799313555
 Attrib cadastroprimeiracompra=A -1.2387056629825806
 Attrib cadastroprimeiracompra=B -5.364019530419439
 Attrib cadastroprimeiracompra=C 7.45917894554552
 Attrib primeiraultimacompra=A 8.252480782751999
 Attrib primeiraultimacompra=C -3.094025734593829

Attrib primeiraultimacompra=B -4.294823273451829

Sigmoid Node 5

Inputs Weights

Threshold -0.8970835013221066

Attrib tipopessoa=F -0.3768268038560354

Attrib tipopessoa=J 0.9603960779084921

Attrib tipopessoa=I 0.28270736222838444

Attrib estado=PR 4.012134518124417

Attrib estado=SP 0.8252279216889982

Attrib estado=MG -0.04767319014068932

Attrib estado=ES 0.6053718166121321

Attrib estado=SC 0.45477061738907837

Attrib estado=RS -2.048981533541426

Attrib estado=RJ -0.2597557257227371

Attrib estado=PI 0.7711110655362146

Attrib estado=GO 0.0207560131160083

Attrib estado=CE 1.6830836523686858

Attrib estado=SE -0.04694457887100121

Attrib estado=DF -0.1499270577688264

Attrib estado=PE -0.33315852387884126

Attrib estado=BA -0.3378808761318507

Attrib estado=AM 0.3398370385299964

Attrib estado=AL 0.1880548220482477

Attrib estado=MT 0.8188607628935914

Attrib estado=MS -0.006261718902453542

Attrib estado=PB -0.22699229236897392

Attrib estado=TO 0.33689125553539795

Attrib estado=MA 0.30206286646561337

Attrib estado=PA 0.31009860549242085

Attrib estado=RN 0.26842528749369976

Attrib estado=RO 0.3101675823518452

Attrib estado=AC 0.42911283485759644

Attrib estado=RR 0.38184066051566434

Attrib estado=AP 0.3810884965799399

Attrib estado=Sp 0.40242521765155614

Attrib estado=sp 0.37079761024604113

Attrib estado=mg 0.3315410299976841

Attrib estado=pr 0.4077240076141888

Attrib estado=Es 0.3385602117388396

Attrib newsletter 1.4148675348825022

Attrib valorgasto=C 0.8446834233733551

Attrib valorgasto=B 0.6304280553514691

Attrib valorgasto=A -0.6672313885947182

Attrib pedidos=A 1.0226073795612238

Attrib pedidos=B -0.8706934634435067

Attrib pedidos=C 0.755266561756906

Attrib taxaaprovacao=C -0.05823111320346754

Attrib taxaaprovacao=B -1.49643147349811

Attrib taxaaprovacao=A 2.460378951474964

Attrib cadastroprimeiracompra=A -3.152311528870408

Attrib cadastroprimeiracompra=B -0.057634805500330746

Attrib cadastroprimeiracompra=C 4.061655469276982

Attrib primeiraultimacompra=A -1.4420730158330357

Attrib primeiraultimacompra=C 0.12035624617250892

Attrib primeiraultimacompra=B 2.2338255128959204

Sigmoid Node 6

Inputs Weights

Threshold -1.4641839923608306

Attrib tipopessoa=F 1.5272980606461133

Attrib tipopessoa=J -2.315334900098211

Attrib tipopessoa=I 2.284886343653162
 Attrib estado=PR -0.4782792468587908
 Attrib estado=SP 1.1690805921910952
 Attrib estado=MG 1.333746993168587
 Attrib estado=ES 0.6849202018589877
 Attrib estado=SC -3.0573114662957472
 Attrib estado=RS 9.542796451049638
 Attrib estado=RJ -3.301915892410527
 Attrib estado=PI -0.2356580355631846
 Attrib estado=GO 4.59958129451414
 Attrib estado=CE -1.5206028237751026
 Attrib estado=SE -2.8436399193923223
 Attrib estado=DF 1.5343053332880996
 Attrib estado=PE 3.5927256659670403
 Attrib estado=BA -2.184534111286251
 Attrib estado=AM 1.3440434005281656
 Attrib estado=AL 8.271056065809251
 Attrib estado=MT 3.3050531537323855
 Attrib estado=MS 0.28364709487325684
 Attrib estado=PB 3.098743150655009
 Attrib estado=TO 1.248268193926677
 Attrib estado=MA 1.6837250437645241
 Attrib estado=PA 1.203577355183019
 Attrib estado=RN -2.5328132216883716
 Attrib estado=RO 1.2687265378776327
 Attrib estado=AC 1.2517701566365573
 Attrib estado=RR 1.261139328919385
 Attrib estado=AP 1.3055929507868087
 Attrib estado=Sp 1.3045447518518298
 Attrib estado=sp 1.303726603086264
 Attrib estado=mg 1.2610631939761074
 Attrib estado=pr 1.284654771066635
 Attrib estado=Es 1.2442120908059475
 Attrib newsletter 3.2799382767550376
 Attrib valorgasto=C 0.30352869720177045
 Attrib valorgasto=B -0.11428868159096665
 Attrib valorgasto=A 1.3342332257325051
 Attrib pedidos=A 3.2857556698849955
 Attrib pedidos=B -0.9134728934210793
 Attrib pedidos=C -0.9226292403293473
 Attrib taxaaprovacao=C 5.936055183981711
 Attrib taxaaprovacao=B -4.8276409860901
 Attrib taxaaprovacao=A 0.3757633984664479
 Attrib cadastroprimeiracompra=A 4.194698937416707
 Attrib cadastroprimeiracompra=B -0.6796493384979729
 Attrib cadastroprimeiracompra=C -1.9478494040624519
 Attrib primeiraultimacompra=A -5.435572957781552
 Attrib primeiraultimacompra=C 2.1461388278929334
 Attrib primeiraultimacompra=B 4.772702661837868

Sigmoid Node 7

Inputs Weights
 Threshold -0.5746191737213576
 Attrib tipopessoa=F -2.7257166514405946
 Attrib tipopessoa=J 2.9470466793969146
 Attrib tipopessoa=I 0.325632721472621
 Attrib estado=PR 2.5905875639460176
 Attrib estado=SP -2.026371741525951
 Attrib estado=MG -1.4159392092985796
 Attrib estado=ES -1.0796494918224937
 Attrib estado=SC 1.5248084777633748

Attrib estado=RS -0.7577605944412195
 Attrib estado=RJ 4.797818060346998
 Attrib estado=PI 1.125895408868033
 Attrib estado=GO -0.34674449132759966
 Attrib estado=CE 0.26863527003254606
 Attrib estado=SE -0.20938059757554492
 Attrib estado=DF -0.15268068490724626
 Attrib estado=PE -0.9419589467019751
 Attrib estado=BA -1.1638703834627557
 Attrib estado=AM 0.1670305426521833
 Attrib estado=AL 0.09992420745006986
 Attrib estado=MT 0.5672479989863601
 Attrib estado=MS 0.5204029659333868
 Attrib estado=PB 0.7140068486077349
 Attrib estado=TO 0.31769524572132857
 Attrib estado=MA 0.07633832288374463
 Attrib estado=PA 0.12407953000188561
 Attrib estado=RN 0.23584615680656984
 Attrib estado=RO 0.2208610972681784
 Attrib estado=AC 0.4664045101387272
 Attrib estado=RR 0.22549132866404156
 Attrib estado=AP 0.2895500907899457
 Attrib estado=Sp 0.2925208717549495
 Attrib estado=sp 0.30552841601341757
 Attrib estado=mg 0.26301970648423517
 Attrib estado=pr 0.25090081122132124
 Attrib estado=Es 0.22950905342855157
 Attrib newsletter -1.0789579897581247
 Attrib valorgasto=C 0.5702388081930226
 Attrib valorgasto=B -0.12918481060774534
 Attrib valorgasto=A 0.08458177821138664
 Attrib pedidos=A -0.0753416416928157
 Attrib pedidos=B 0.20405895501616872
 Attrib pedidos=C 0.4205717862985689
 Attrib taxaaprovacao=C -0.6025576189080344
 Attrib taxaaprovacao=B 3.2317101577195992
 Attrib taxaaprovacao=A -2.030646806473484
 Attrib cadastroprimeiracompra=A -5.702632866770078
 Attrib cadastroprimeiracompra=B 3.480221665828944
 Attrib cadastroprimeiracompra=C 2.8245964131754087
 Attrib primeiraultimacompra=A -0.4010214813791953
 Attrib primeiraultimacompra=C -1.038063036861205
 Attrib primeiraultimacompra=B 2.0517137546417406

Sigmoid Node 8

Inputs Weights

Threshold -0.7555532678972557
 Attrib tipopessoa=F -4.1020093091886025
 Attrib tipopessoa=J 4.708466548879592
 Attrib tipopessoa=I 0.13157109599479666
 Attrib estado=PR 1.743366451187803
 Attrib estado=SP -2.628155230435138
 Attrib estado=MG 4.870024035972265
 Attrib estado=ES 0.36636547919272405
 Attrib estado=SC 0.8659597202673337
 Attrib estado=RS 0.2842337221368304
 Attrib estado=RJ -1.7808041949840434
 Attrib estado=PI 0.20577322325658093
 Attrib estado=GO 0.010134852059815437
 Attrib estado=CE 0.3974889898378179
 Attrib estado=SE -0.7928849065516201

Attrib estado=DF -0.009182046071573576
 Attrib estado=PE 0.01674671531859504
 Attrib estado=BA -0.09398745067313835
 Attrib estado=AM 0.2798067408301369
 Attrib estado=AL 0.30537867476528385
 Attrib estado=MT 0.024147688868374102
 Attrib estado=MS 0.3328708441402321
 Attrib estado=PB 0.49169378087088145
 Attrib estado=TO 0.41976376362359086
 Attrib estado=MA 0.19545539936179054
 Attrib estado=PA 0.22609777395347705
 Attrib estado=RN 0.21720183486776842
 Attrib estado=RO 0.32746323507999825
 Attrib estado=AC 0.33222281922718083
 Attrib estado=RR 0.32471354129079416
 Attrib estado=AP 0.3056404870706811
 Attrib estado=Sp 0.32176855054386067
 Attrib estado=sp 0.24332139001728378
 Attrib estado=mg 0.3246075179397668
 Attrib estado=pr 0.31300915985233685
 Attrib estado=Es 0.25260477851051116
 Attrib newsletter -1.7164681084946636
 Attrib valorgasto=C 0.783940689208638
 Attrib valorgasto=B 0.36296223965604635
 Attrib valorgasto=A -0.3584638929168899
 Attrib pedidos=A -0.5875158835103993
 Attrib pedidos=B 0.7238974376313402
 Attrib pedidos=C 0.5967187723991909
 Attrib taxaaprovacao=C -0.28446440225108244
 Attrib taxaaprovacao=B -0.25830646905137694
 Attrib taxaaprovacao=A 1.2886214642216
 Attrib cadastroprimeiracompra=A -3.6572866975859406
 Attrib cadastroprimeiracompra=B 1.1049709611869492
 Attrib cadastroprimeiracompra=C 3.3657002552108573
 Attrib primeiraultimacompra=A -1.0941663612889496
 Attrib primeiraultimacompra=C 2.3995707172193885
 Attrib primeiraultimacompra=B -0.538025592742646

Sigmoid Node 9

Inputs Weights

Threshold 1.6215553142226757
 Attrib tipopessoa=F -3.29900676321593
 Attrib tipopessoa=J 5.113160107416251
 Attrib tipopessoa=I -3.4248613433575903
 Attrib estado=PR -3.643266767369724
 Attrib estado=SP -0.7653547906764404
 Attrib estado=MG -3.5300432736583636
 Attrib estado=ES -0.055256170926307406
 Attrib estado=SC 0.2013323708051856
 Attrib estado=RS -2.8481237560143526
 Attrib estado=RJ -1.1129780026322074
 Attrib estado=PI 0.6986419114128366
 Attrib estado=GO 4.636338448407783
 Attrib estado=CE 2.469122575849789
 Attrib estado=SE 5.189437115269178
 Attrib estado=DF 3.41607032035336
 Attrib estado=PE 1.7964676508483879
 Attrib estado=BA -0.07733031721045854
 Attrib estado=AM 0.7319304514409943
 Attrib estado=AL 0.6033950817473707
 Attrib estado=MT 2.681276357961521

Attrib estado=MS -0.14553590907034408
 Attrib estado=PB -0.8814065935106733
 Attrib estado=TO 0.46592576042431366
 Attrib estado=MA 2.459321439329244
 Attrib estado=PA 0.8842471836491794
 Attrib estado=RN 2.1446506339585576
 Attrib estado=RO 1.3305966980992565
 Attrib estado=AC 0.6632485640974001
 Attrib estado=RR 0.7401508901827911
 Attrib estado=AP 0.7565424659158234
 Attrib estado=Sp 0.750766983785309
 Attrib estado=sp 0.725212223455811
 Attrib estado=mg 0.6942081938024366
 Attrib estado=pr 0.7284191309025674
 Attrib estado=Es 0.7825429046639614
 Attrib newsletter -1.752081960097165
 Attrib valorgasto=C -1.5461592026766744
 Attrib valorgasto=B -0.9797648390380227
 Attrib valorgasto=A 0.9629964292985574
 Attrib pedidos=A 1.555185606382677
 Attrib pedidos=B -1.648728478452314
 Attrib pedidos=C -1.5780834954043022
 Attrib taxaaprovacao=C -3.776490394506592
 Attrib taxaaprovacao=B 0.9137122373640522
 Attrib taxaaprovacao=A 1.2347463944959682
 Attrib cadastroprimeiracompra=A -2.686357194191144
 Attrib cadastroprimeiracompra=B -0.9068456730221227
 Attrib cadastroprimeiracompra=C 2.0083584014047853
 Attrib primeiraultimacompra=A -1.778856458250367
 Attrib primeiraultimacompra=C -1.331813510129199
 Attrib primeiraultimacompra=B 1.4887596921719017

Sigmoid Node 10

Inputs Weights

Threshold -0.8373407323754565
 Attrib tipopessoa=F -2.4148571011907345
 Attrib tipopessoa=J 3.70810725981798
 Attrib tipopessoa=I -0.4802279037279884
 Attrib estado=PR 2.311606259437495
 Attrib estado=SP 0.5932309362103834
 Attrib estado=MG -0.02284810437975952
 Attrib estado=ES -0.2545562960189139
 Attrib estado=SC -0.33006749756106624
 Attrib estado=RS -0.7846631683177829
 Attrib estado=RJ 3.2889073943691196
 Attrib estado=PI 0.36769300650765485
 Attrib estado=GO -0.1478839492503374
 Attrib estado=CE 0.411519857297558
 Attrib estado=SE -0.055555231924000956
 Attrib estado=DF -0.30651563518696895
 Attrib estado=PE -0.3464150665062674
 Attrib estado=BA -0.5037780300610696
 Attrib estado=AM 0.19727855157449448
 Attrib estado=AL 0.2270594922820194
 Attrib estado=MT 0.27259282449646477
 Attrib estado=MS 0.3364165841688399
 Attrib estado=PB -0.5354323957544713
 Attrib estado=TO 0.7113497493160976
 Attrib estado=MA -0.00545637936458749
 Attrib estado=PA 0.20747346354212315
 Attrib estado=RN 0.08462613231842729

Attrib estado=RO 0.2858138828951001
 Attrib estado=AC 0.40987057249560727
 Attrib estado=RR 0.2597349197314526
 Attrib estado=AP 0.2286097507327742
 Attrib estado=Sp 0.2921945037923607
 Attrib estado=sp 0.27344655387462335
 Attrib estado=mg 0.2767373949116531
 Attrib estado=pr 0.23148074630290708
 Attrib estado=Es 0.2957495750902726
 Attrib newsletter -0.6746805961093274
 Attrib valorgasto=C 0.8568892646680591
 Attrib valorgasto=B 0.21919567803603898
 Attrib valorgasto=A -0.14762711925733407
 Attrib pedidos=A 0.018352416009817105
 Attrib pedidos=B 0.11650061575709542
 Attrib pedidos=C 0.7270948946815007
 Attrib taxaaprovacao=C -0.7469589750813465
 Attrib taxaaprovacao=B 0.616972434130058
 Attrib taxaaprovacao=A 1.0210594893838032
 Attrib cadastroprimeiracompra=A -2.7340238548894367
 Attrib cadastroprimeiracompra=B 0.9226663898381141
 Attrib cadastroprimeiracompra=C 2.7696233327516304
 Attrib primeiraultimacompra=A 0.5737367916549438
 Attrib primeiraultimacompra=C -0.11883766828223749
 Attrib primeiraultimacompra=B 0.41116053928849966

Sigmoid Node 11

Inputs Weights

Threshold -1.1218547699503725

Attrib tipopessoa=F -3.3835886229497762
 Attrib tipopessoa=J 5.829732352993477
 Attrib tipopessoa=I -1.406113038079613
 Attrib estado=PR 1.9407749882529644
 Attrib estado=SP -1.9082763279714725
 Attrib estado=MG 6.127298236982054
 Attrib estado=ES 0.23486204545020853
 Attrib estado=SC -2.018834888251633
 Attrib estado=RS -0.1362286265301833
 Attrib estado=RJ -1.8981911173913157
 Attrib estado=PI 0.2073727346225657
 Attrib estado=GO 0.03183076959356301
 Attrib estado=CE 0.6912695721451071
 Attrib estado=SE -0.028036103734863566
 Attrib estado=DF 0.059956055127449455
 Attrib estado=PE 0.036321749162254426
 Attrib estado=BA -0.2318139968635907
 Attrib estado=AM 0.28651358724009607
 Attrib estado=AL 0.26024086509616806
 Attrib estado=MT 0.26789160538355666
 Attrib estado=MS 0.3711271824101972
 Attrib estado=PB 0.5189269307011967
 Attrib estado=TO 0.4272802512856615
 Attrib estado=MA 0.17892174200974234
 Attrib estado=PA 0.22465257743883005
 Attrib estado=RN 0.30290334937106694
 Attrib estado=RO 0.34357472368187075
 Attrib estado=AC 0.32824532018568103
 Attrib estado=RR 0.22437593858914645
 Attrib estado=AP 0.2813783789638379
 Attrib estado=Sp 0.3312915329014293
 Attrib estado=sp 0.29301925490536274

Attrib estado=mg 0.26899971382209753
 Attrib estado=pr 0.248724155621682
 Attrib estado=Es 0.3287187357050839
 Attrib newsletter -0.30511587035620635
 Attrib valorgasto=C 1.1018595077454916
 Attrib valorgasto=B 0.03639280046999948
 Attrib valorgasto=A -0.014368821516021655
 Attrib pedidos=A -1.0207059914907093
 Attrib pedidos=B 1.1584367712509451
 Attrib pedidos=C 0.9944910605969001
 Attrib taxaaprovacao=C -1.82860223003348
 Attrib taxaaprovacao=B 0.6888882466939212
 Attrib taxaaprovacao=A 2.2898440805346123
 Attrib cadastroprimeiracompra=A -0.799451651065178
 Attrib cadastroprimeiracompra=B -0.3721839470044083
 Attrib cadastroprimeiracompra=C 2.1393896595898543
 Attrib primeiraultimacompra=A -1.4569376848283313
 Attrib primeiraultimacompra=C 0.8901168400425522
 Attrib primeiraultimacompra=B 1.6858965877392575

Sigmoid Node 12

Inputs Weights

Threshold -0.8607608676330137
 Attrib tipopessoa=F -0.9990921219330859
 Attrib tipopessoa=J 2.873157120935851
 Attrib tipopessoa=I -0.972179219483567
 Attrib estado=PR -0.9142485739767525
 Attrib estado=SP -1.7416994790325424
 Attrib estado=MG -1.6976243551201704
 Attrib estado=ES 2.940883391059905
 Attrib estado=SC 0.21181032262923294
 Attrib estado=RS -1.6361218177225696
 Attrib estado=RJ -1.5468088168127472
 Attrib estado=PI -0.5763831815151355
 Attrib estado=GO -1.1214135788241362
 Attrib estado=CE 0.5357690611063204
 Attrib estado=SE -0.9720887608863619
 Attrib estado=DF 7.456659995774459
 Attrib estado=PE -0.563939197298257
 Attrib estado=BA -0.5183066807065072
 Attrib estado=AM -0.4997516878641718
 Attrib estado=AL -1.169935271427652
 Attrib estado=MT 3.33944759398691
 Attrib estado=MS -0.7784561844761785
 Attrib estado=PB 2.762323341113929
 Attrib estado=TO -0.3744169652111761
 Attrib estado=MA -0.3545378031418724
 Attrib estado=PA 0.21518601896226866
 Attrib estado=RN 3.9420676834613757
 Attrib estado=RO -1.4513827068942733
 Attrib estado=AC 7.867387609354754
 Attrib estado=RR -0.05554792708000198
 Attrib estado=AP 0.23628720587017496
 Attrib estado=Sp 0.49934045162351115
 Attrib estado=sp 0.5827649235297245
 Attrib estado=mg 0.5544560422115445
 Attrib estado=pr 0.5855149717326003
 Attrib estado=Es 0.5482120112114283
 Attrib newsletter 0.7939545704611211
 Attrib valorgasto=C 0.9991016211855281
 Attrib valorgasto=B 0.47373136862868837

Attrib valorgasto=A -0.5163302885265262
Attrib pedidos=A -0.5263183833467411
Attrib pedidos=B 0.725184805233572
Attrib pedidos=C 0.6597177839725216
Attrib taxaaprovacao=C -0.1725130618153727
Attrib taxaaprovacao=B -0.43350337629025787
Attrib taxaaprovacao=A 1.4969467812954087
Attrib cadastroprimeiracompra=A 2.829911463305982
Attrib cadastroprimeiracompra=B -2.73084397688083
Attrib cadastroprimeiracompra=C 0.7309810799377116
Attrib primeiraultimacompra=A 2.2321590963363107
Attrib primeiraultimacompra=C 0.5835114717537967
Attrib primeiraultimacompra=B -1.8860823778509834

Sigmoid Node 13

Inputs Weights

Threshold -0.769247736645894

Attrib tipopessoa=F -0.8355547822674548
Attrib tipopessoa=J 2.8232350076513404
Attrib tipopessoa=I -1.1856253875503395
Attrib estado=PR 3.010209724666501
Attrib estado=SP 0.9967770529556628
Attrib estado=MG -0.9000631341769222
Attrib estado=ES -0.7720784489114438
Attrib estado=SC -1.1991628797812237
Attrib estado=RS -0.3969658976700327
Attrib estado=RJ -0.40857880105530425
Attrib estado=PI 0.3068422229328082
Attrib estado=GO -0.07923558527466883
Attrib estado=CE 0.07780591207520324
Attrib estado=SE -0.04020858700391054
Attrib estado=DF -0.33951571386003165
Attrib estado=PE -0.32101855021191067
Attrib estado=BA 1.1723683439984391
Attrib estado=AM 0.26730296975393686
Attrib estado=AL 0.17932464802601913
Attrib estado=MT -0.1754062241407095
Attrib estado=MS 0.37275903523179116
Attrib estado=PB 3.5478394197817407
Attrib estado=TO 0.7404840772535278
Attrib estado=MA -0.138663556488786
Attrib estado=PA 0.16437159256905473
Attrib estado=RN -0.01052106575682052
Attrib estado=RO -0.1424442832961975
Attrib estado=AC 0.37004667245254025
Attrib estado=RR 0.28843702019530587
Attrib estado=AP 0.26484472606707604
Attrib estado=Sp 0.22777823930567517
Attrib estado=sp 0.2522125875683197
Attrib estado=mg 0.23808894366784591
Attrib estado=pr 0.2556271107956544
Attrib estado=Es 0.25068608775797896
Attrib newsletter 0.7071060653439966
Attrib valorgasto=C 0.8074099669243037
Attrib valorgasto=B 0.6638311908707871
Attrib valorgasto=A -0.6895566982817963
Attrib pedidos=A -1.9828085222491634
Attrib pedidos=B 2.102624407692184
Attrib pedidos=C 0.6721658209646287
Attrib taxaaprovacao=C -1.1030845145518222
Attrib taxaaprovacao=B -1.1548453895535467

Attrib taxaaprovacao=A 3.059641587251259
Attrib cadastroprimeiracompra=A 1.352450842807729
Attrib cadastroprimeiracompra=B -0.44654347431256636
Attrib cadastroprimeiracompra=C -0.1635264035535638
Attrib primeiraultimacompra=A -0.1874540519416809
Attrib primeiraultimacompra=C -0.2585194398102394
Attrib primeiraultimacompra=B 1.186445108552489

Sigmoid Node 14

Inputs *Weights*
Threshold -3.069238458747206
Attrib tipopessoa=F -1.129379378540063
Attrib tipopessoa=J 3.4142280577750497
Attrib tipopessoa=I 0.8581041943901881
Attrib estado=PR 0.4237068627625012
Attrib estado=SP -3.1572148766566817
Attrib estado=MG -2.0618702986421655
Attrib estado=ES -0.41418355926035655
Attrib estado=SC 0.6409084304872698
Attrib estado=RS -1.6831937215814852
Attrib estado=RJ -2.6759705659376603
Attrib estado=PI -0.28577652616731214
Attrib estado=GO -0.20112007324448528
Attrib estado=CE 2.2091518727514416
Attrib estado=SE -0.3369043849568489
Attrib estado=DF 0.4153435588764572
Attrib estado=PE -0.24225422458246018
Attrib estado=BA 1.9500433221156943
Attrib estado=AM 0.07794014529701204
Attrib estado=AL -0.11036622010625732
Attrib estado=MT -1.550801230953772
Attrib estado=MS 0.3151042068999551
Attrib estado=PB 1.1396680840575841
Attrib estado=TO 4.552163355255775
Attrib estado=MA 2.767665533635869
Attrib estado=PA 0.019244159777696612
Attrib estado=RN -0.15539605134755607
Attrib estado=RO -0.07181942098162405
Attrib estado=AC 0.2843509941844461
Attrib estado=RR 0.012359770661305497
Attrib estado=AP 0.06514723474639476
Attrib estado=Sp 0.08003015081113016
Attrib estado=sp 0.04792592888614121
Attrib estado=mg 0.06289541030701398
Attrib estado=pr 0.07074956433117892
Attrib estado=Es 0.0681781202557007
Attrib newsletter -12.60645204872489
Attrib valorgasto=C 1.9725160693416819
Attrib valorgasto=B 2.6082093686479912
Attrib valorgasto=A -1.5011583562193551
Attrib pedidos=A -1.6643400364860834
Attrib pedidos=B 1.9002879482520951
Attrib pedidos=C 2.953439397694719
Attrib taxaaprovacao=C 3.7450098913251035
Attrib taxaaprovacao=B -2.334403550408135
Attrib taxaaprovacao=A 1.7296445341003477
Attrib cadastroprimeiracompra=A 6.82573923598184
Attrib cadastroprimeiracompra=B -2.806258711830728
Attrib cadastroprimeiracompra=C -0.9113119279908827
Attrib primeiraultimacompra=A 5.40726465138781
Attrib primeiraultimacompra=C -1.1181492805789066

Attrib primeiraultimacompra=B -1.226776772222842
 Sigmoid Node 15
Inputs Weights
Threshold -0.026154041066895173
Attrib tipopessoa=F -1.6307817170250292
Attrib tipopessoa=J -0.7237014482326268
Attrib tipopessoa=I 2.4224967875253367
Attrib estado=PR -0.9329456049627339
Attrib estado=SP -1.8099937009374656
Attrib estado=MG 2.537195927303307
Attrib estado=ES 3.0542608847941604
Attrib estado=SC -1.119254350989464
Attrib estado=RS 3.057288426023129
Attrib estado=RJ -0.8589782970110176
Attrib estado=PI 0.33174176253092713
Attrib estado=GO 0.07976773130064849
Attrib estado=CE 0.3558751543669662
Attrib estado=SE 0.05683955278016154
Attrib estado=DF 0.33054910462531806
Attrib estado=PE 2.037162858182506
Attrib estado=BA 0.6328071964087407
Attrib estado=AM 0.49906613649403053
Attrib estado=AL 0.11272267811942274
Attrib estado=MT 0.16360035783256027
Attrib estado=MS 0.47220039888891685
Attrib estado=PB 1.2114553533801602
Attrib estado=TO 0.5141788736852888
Attrib estado=MA -1.0995582590966486
Attrib estado=PA 0.7353283640787952
Attrib estado=RN 0.3350081598070984
Attrib estado=RO 0.498550400466481
Attrib estado=AC 0.543024419309938
Attrib estado=RR 0.48773931805960286
Attrib estado=AP 0.5262681286958534
Attrib estado=Sp 0.48323320874963055
Attrib estado=sp 0.5325846176199246
Attrib estado=mg 0.5369023617522719
Attrib estado=pr 0.4973145024939567
Attrib estado=Es 0.5226089499292716
Attrib newsletter 0.6356596701640177
Attrib valorgasto=C 0.020619979591716877
Attrib valorgasto=B -0.2642312733067849
Attrib valorgasto=A 0.2615019536865861
Attrib pedidos=A 0.6765514577332258
Attrib pedidos=B -0.2013534199059428
Attrib pedidos=C -0.403205599959876
Attrib taxaaprovacao=C 2.4772337265487545
Attrib taxaaprovacao=B -1.4395976106039925
Attrib taxaaprovacao=A -0.9650795709849078
Attrib cadastroprimeiracompra=A -2.2991565140163552
Attrib cadastroprimeiracompra=B 3.6732075138686713
Attrib cadastroprimeiracompra=C -1.3090888132677798
Attrib primeiraultimacompra=A -1.4253068521773813
Attrib primeiraultimacompra=C 1.218275803633339
Attrib primeiraultimacompra=B 0.1953728606228364
 Sigmoid Node 16
Inputs Weights
Threshold -0.551967124668439
Attrib tipopessoa=F -0.5747836143917604
Attrib tipopessoa=J -0.8550445711974674

Attrib tipopessoa=I 2.0618598742348837
Attrib estado=PR -0.44231197927274074
Attrib estado=SP -2.221218739281752
Attrib estado=MG 0.43557558579323796
Attrib estado=ES 1.6628108526932457
Attrib estado=SC -1.3071767131045813
Attrib estado=RS -0.5295715542369261
Attrib estado=RJ -1.473756087525149
Attrib estado=PI 0.17364294590102308
Attrib estado=GO 1.2000587191565701
Attrib estado=CE -0.7508699755451401
Attrib estado=SE -0.9078910600386453
Attrib estado=DF -0.6529651290982121
Attrib estado=PE 1.7447984029946815
Attrib estado=BA 1.587448881540653
Attrib estado=AM 0.7564776096259267
Attrib estado=AL 5.022686789681207
Attrib estado=MT -0.7424526695620458
Attrib estado=MS 0.2844449761969487
Attrib estado=PB 4.157221529843363
Attrib estado=TO 0.7426481287767602
Attrib estado=MA 5.13992345242805
Attrib estado=PA 0.7821653268775307
Attrib estado=RN -0.127610962535544
Attrib estado=RO 0.6723127182839469
Attrib estado=AC 0.7470132099689464
Attrib estado=RR 0.7275428759843484
Attrib estado=AP 0.7452910788189011
Attrib estado=Sp 0.7295496094172824
Attrib estado=sp 0.6656319542857179
Attrib estado=mg 0.7396415618308165
Attrib estado=pr 0.7407207305668342
Attrib estado=Es 0.6764143247390988
Attrib newsletter 0.7819232586251591
Attrib valorgasto=C 0.533331106210623
Attrib valorgasto=B 0.42606280255683127
Attrib valorgasto=A -0.4669848456826354
Attrib pedidos=A 0.39360328982255016
Attrib pedidos=B 0.38073555616448784
Attrib pedidos=C -0.26456928742684993
Attrib taxaaprovacao=C 3.4532473413886158
Attrib taxaaprovacao=B -0.6048482022645553
Attrib taxaaprovacao=A -2.1751610632612652
Attrib cadastroprimeiracompra=A 2.7216997598451274
Attrib cadastroprimeiracompra=B -1.1944849309413776
Attrib cadastroprimeiracompra=C -0.9531041859488327
Attrib primeiraultimacompra=A -1.2210607054114468
Attrib primeiraultimacompra=C -1.0774114545258584
Attrib primeiraultimacompra=B 2.896031565031605

Sigmoid Node 17

Inputs *Weights*
Threshold -0.0653348784717629
Attrib tipopessoa=F -1.8378413795302546
Attrib tipopessoa=J -0.8181198868871249
Attrib tipopessoa=I 2.675241202622706
Attrib estado=PR 0.37449935347138014
Attrib estado=SP -0.16800394796310872
Attrib estado=MG 0.3542318271319513
Attrib estado=ES 0.34198132728067254
Attrib estado=SC 0.44697064409117737

Attrib estado=RS 0.5741830784669347
 Attrib estado=RJ 0.18535273349367587
 Attrib estado=PI 0.47271633786798484
 Attrib estado=GO 0.27549589223903226
 Attrib estado=CE 0.563465827775448
 Attrib estado=SE 0.3456599741174149
 Attrib estado=DF 0.5520310858638365
 Attrib estado=PE 0.4236940037089832
 Attrib estado=BA 0.5057402930865523
 Attrib estado=AM 0.48765159299763416
 Attrib estado=AL 0.38560316168244957
 Attrib estado=MT 0.3976959782736411
 Attrib estado=MS 0.40901627932200235
 Attrib estado=PB 0.4403305546980491
 Attrib estado=TO 0.4705566066545277
 Attrib estado=MA 0.485168650107583
 Attrib estado=PA 0.39044677198116
 Attrib estado=RN 0.4865730326716186
 Attrib estado=RO 0.4394950813474863
 Attrib estado=AC 0.4059626328999168
 Attrib estado=RR 0.46680353868469926
 Attrib estado=AP 0.4720138629018873
 Attrib estado=Sp 0.42234251313815074
 Attrib estado=sp 0.46508113607953716
 Attrib estado=mg 0.46280744245220623
 Attrib estado=pr 0.424792791390753
 Attrib estado=Es 0.44006140004044153
 Attrib newsletter -0.8135980137177954
 Attrib valorgasto=C 0.036360882596283395
 Attrib valorgasto=B -0.028187285231655618
 Attrib valorgasto=A 0.07981145204634817
 Attrib pedidos=A 0.03320338722032836
 Attrib pedidos=B -0.010886396426596045
 Attrib pedidos=C -0.0312965874661073
 Attrib taxaaprovacao=C 0.6931525062704238
 Attrib taxaaprovacao=B -0.19854554822962855
 Attrib taxaaprovacao=A -0.44582620967098024
 Attrib cadastroprimeiracompra=A -0.1638300352702192
 Attrib cadastroprimeiracompra=B -0.08037329523387499
 Attrib cadastroprimeiracompra=C 0.24965737804435467
 Attrib primeiraultimacompra=A -0.24049932701622026
 Attrib primeiraultimacompra=C 0.41699778012166344
 Attrib primeiraultimacompra=B -0.1509420570441154

Sigmoid Node 18

Inputs Weights

Threshold -1.9032518055186542
 Attrib tipopessoa=F 1.4289103816672397
 Attrib tipopessoa=J 1.8483519991846658
 Attrib tipopessoa=I -1.4490454886580133
 Attrib estado=PR 1.8083215264665309
 Attrib estado=SP -2.990353386976917
 Attrib estado=MG -2.243636906638274
 Attrib estado=ES 4.4678528451733355
 Attrib estado=SC 2.3747539969704383
 Attrib estado=RS -2.4474815865371666
 Attrib estado=RJ -1.5766064126443389
 Attrib estado=PI -0.5848141375758754
 Attrib estado=GO -0.11357400496734223
 Attrib estado=CE 6.4875158442616305
 Attrib estado=SE 0.8216292881067653

Attrib estado=DF -1.524506878004072
 Attrib estado=PE 0.023050134954864847
 Attrib estado=BA -2.3557101055527574
 Attrib estado=AM -0.45775723947034125
 Attrib estado=AL -0.7980630538452839
 Attrib estado=MT 5.043026233196275
 Attrib estado=MS -1.6983005211565982
 Attrib estado=PB 6.311464864738352
 Attrib estado=TO -0.01624033724126007
 Attrib estado=MA -1.159220559293839
 Attrib estado=PA -0.07785961200964293
 Attrib estado=RN -1.851284273870014
 Attrib estado=RO -1.5165481907149005
 Attrib estado=AC 6.310777996091333
 Attrib estado=RR -0.6342261367809677
 Attrib estado=AP 0.2915448318481098
 Attrib estado=Sp 0.5161960883818665
 Attrib estado=sp 0.46087240623251663
 Attrib estado=mg 0.46121068481550265
 Attrib estado=pr 0.48231029128198205
 Attrib estado=Es 0.506650943172679
 Attrib newsletter -2.997169854921904
 Attrib valorgasto=C 1.222712698947741
 Attrib valorgasto=B 0.6429180382738439
 Attrib valorgasto=A 0.13037644499621973
 Attrib pedidos=A -0.9707417674231229
 Attrib pedidos=B 1.1020027396584091
 Attrib pedidos=C 1.7387353480306393
 Attrib taxaaprovacao=C 0.13854923366739122
 Attrib taxaaprovacao=B -1.6200729635493547
 Attrib taxaaprovacao=A 3.346893164754114
 Attrib cadastroprimeiracompra=A 3.3046400269809757
 Attrib cadastroprimeiracompra=B -2.9409080377909795
 Attrib cadastroprimeiracompra=C 1.5482904650601423
 Attrib primeiraultimacompra=A 4.926482838139154
 Attrib primeiraultimacompra=C -1.56532244826725
 Attrib primeiraultimacompra=B -1.369200828053641

Sigmoid Node 19

Inputs Weights
 Threshold -1.789370626745014
 Attrib tipopessoa=F -0.4772230161534432
 Attrib tipopessoa=J 1.5690289930564172
 Attrib tipopessoa=I 0.7286625421380529
 Attrib estado=PR -0.46688681011120603
 Attrib estado=SP -2.004573267865076
 Attrib estado=MG -1.4375019615043096
 Attrib estado=ES -0.6902008994594266
 Attrib estado=SC -1.4201516044859337
 Attrib estado=RS -1.6789719547235389
 Attrib estado=RJ 3.55697849216593
 Attrib estado=PI 0.42439986849669953
 Attrib estado=GO 0.6255375415763306
 Attrib estado=CE 0.28160746755032395
 Attrib estado=SE 0.10341053798312665
 Attrib estado=DF 3.7822001603502042
 Attrib estado=PE -0.19243607413948774
 Attrib estado=BA 0.5013061704381554
 Attrib estado=AM 0.11816897107246724
 Attrib estado=AL -0.06297641076315387
 Attrib estado=MT 0.3736900938135062

Attrib estado=MS -0.4590562155554455
 Attrib estado=PB 0.2867586331126133
 Attrib estado=TO 0.09049244278001538
 Attrib estado=MA 0.07938919821028308
 Attrib estado=PA 0.06815682474668121
 Attrib estado=RN 0.1830513272942696
 Attrib estado=RO 0.15908746968462048
 Attrib estado=AC 0.15994465576010086
 Attrib estado=RR 0.1393738306910534
 Attrib estado=AP 0.08252847124275167
 Attrib estado=Sp 0.11824735649167005
 Attrib estado=sp 0.1404106531796913
 Attrib estado=mg 0.09312077941237992
 Attrib estado=pr 0.1212188728630541
 Attrib estado=Es 0.0840888022567535
 Attrib newsletter -7.36044415366086
 Attrib valorgasto=C 1.7998990849510739
 Attrib valorgasto=B 0.9349416811800367
 Attrib valorgasto=A -0.8875723553394748
 Attrib pedidos=A -1.6493543061530933
 Attrib pedidos=B 1.7461101935780245
 Attrib pedidos=C 1.82196879253474
 Attrib taxaaprovacao=C 0.4529985479881114
 Attrib taxaaprovacao=B -0.05710858961504088
 Attrib taxaaprovacao=A 1.381440475759475
 Attrib cadastroprimeiracompra=A 1.7671992282557953
 Attrib cadastroprimeiracompra=B -2.2560741227068393
 Attrib cadastroprimeiracompra=C 2.286888863863332
 Attrib primeiraultimacompra=A 2.158619786449828
 Attrib primeiraultimacompra=C -2.2204755257439315
 Attrib primeiraultimacompra=B 1.920466209403556

Sigmoid Node 20

Inputs Weights
 Threshold -1.6392583310387228
 Attrib tipopessoa=F -1.1499258085694666
 Attrib tipopessoa=J 0.9015566504469005
 Attrib tipopessoa=I 1.9452975099275396
 Attrib estado=PR -2.1974214042691527
 Attrib estado=SP -2.1297510480016792
 Attrib estado=MG -0.285462405301079
 Attrib estado=ES 4.495439736900838
 Attrib estado=SC 2.084952529497589
 Attrib estado=RS -2.9708429925667077
 Attrib estado=RJ 0.6553585055317522
 Attrib estado=PI 5.363405785859758
 Attrib estado=GO 0.1251376708732324
 Attrib estado=CE 1.6570967288185727
 Attrib estado=SE -0.007277477128511319
 Attrib estado=DF 1.249573155097708
 Attrib estado=PE -0.12681394531829182
 Attrib estado=BA 0.14148761859586007
 Attrib estado=AM 0.5707564894920557
 Attrib estado=AL 0.049215846247402745
 Attrib estado=MT 0.024498962442998175
 Attrib estado=MS -3.2499466997530444
 Attrib estado=PB 0.8002152481433878
 Attrib estado=TO 0.14036567648653017
 Attrib estado=MA 2.845012185088701
 Attrib estado=PA 0.5262851382471674
 Attrib estado=RN 0.446489482596466

Attrib estado=RO 0.7761148598214846
 Attrib estado=AC 0.5942407120703708
 Attrib estado=RR 0.4635261257603966
 Attrib estado=AP 0.5081476625599626
 Attrib estado=Sp 0.5285257850672898
 Attrib estado=sp 0.5594018327118875
 Attrib estado=mg 0.5318979412210156
 Attrib estado=pr 0.4780798242624287
 Attrib estado=Es 0.5072937432339435
 Attrib newsletter -8.299651025053162
 Attrib valorgasto=C 3.22836977918588
 Attrib valorgasto=B -6.478386099886096
 Attrib valorgasto=A 4.876502472229357
 Attrib pedidos=A -5.112313227025962
 Attrib pedidos=B 5.570707516571954
 Attrib pedidos=C 1.2489747538160694
 Attrib taxaaprovacao=C 13.493993196787502
 Attrib taxaaprovacao=B -1.4792955082317527
 Attrib taxaaprovacao=A -10.437025609029707
 Attrib cadastroprimeiracompra=A -2.991061777783762
 Attrib cadastroprimeiracompra=B 1.4934950490881647
 Attrib cadastroprimeiracompra=C 3.2165391416312263
 Attrib primeiraultimacompra=A 12.977208377933728
 Attrib primeiraultimacompra=C -2.1725423254169334
 Attrib primeiraultimacompra=B -9.103063557072641

Sigmoid Node 21

Inputs Weights

Threshold -1.0179705280454472
 Attrib tipopessoa=F 0.3029601006640357
 Attrib tipopessoa=J 0.026035885265508752
 Attrib tipopessoa=I 0.7656238847455509
 Attrib estado=PR 0.7920443810232174
 Attrib estado=SP 2.5117612458397356
 Attrib estado=MG -1.0811877936838221
 Attrib estado=ES 0.30858622329697843
 Attrib estado=SC -0.6766517407540774
 Attrib estado=RS -0.6986259080500622
 Attrib estado=RJ -1.298072604625048
 Attrib estado=PI 0.7424753852504886
 Attrib estado=GO -0.020964746935680956
 Attrib estado=CE 0.5123156103615301
 Attrib estado=SE 0.0514544702579081
 Attrib estado=DF 0.6883355927332262
 Attrib estado=PE -0.1781815188488285
 Attrib estado=BA -0.5057361024710738
 Attrib estado=AM 0.3825269780603473
 Attrib estado=AL 0.12438832345842332
 Attrib estado=MT 5.493810294297533
 Attrib estado=MS -0.2354516016095649
 Attrib estado=PB 0.7747756337177109
 Attrib estado=TO 0.2458798961593329
 Attrib estado=MA 0.32339827938470056
 Attrib estado=PA 0.4716201037207499
 Attrib estado=RN 0.4891624365074202
 Attrib estado=RO 0.31932537483456597
 Attrib estado=AC 0.5169890860937472
 Attrib estado=RR 0.39966234207643503
 Attrib estado=AP 0.4320016853558824
 Attrib estado=Sp 0.39853433601735716
 Attrib estado=sp 0.39186108129468816

Attrib estado=mg 0.4193529758038636
 Attrib estado=pr 0.43920879167949645
 Attrib estado=Es 0.4662593939500747
 Attrib newsletter -3.210858200863663
 Attrib valorgasto=C 0.9876046532141542
 Attrib valorgasto=B 0.008857111419244479
 Attrib valorgasto=A 0.08105549122235949
 Attrib pedidos=A -0.638642088705357
 Attrib pedidos=B 0.6979026665601861
 Attrib pedidos=C 0.9955289448827334
 Attrib taxaaprovacao=C -1.0945001260451728
 Attrib taxaaprovacao=B -0.40548318377844283
 Attrib taxaaprovacao=A 2.4939299558043375
 Attrib cadastroprimeiracompra=A -1.3820548105770631
 Attrib cadastroprimeiracompra=B -0.7111205628635244
 Attrib cadastroprimeiracompra=C 3.0756738243662403
 Attrib primeiraultimacompra=A 1.9739251952418515
 Attrib primeiraultimacompra=C -0.12872606890804647
 Attrib primeiraultimacompra=B -0.8228997966879534

Sigmoid Node 22

Inputs Weights

Threshold -2.5778551509505183
 Attrib tipopessoa=F 0.7774154722455193
 Attrib tipopessoa=J -2.9079413506103506
 Attrib tipopessoa=I 4.676150769584396
 Attrib estado=PR -1.8978689463762428
 Attrib estado=SP -2.160695205197505
 Attrib estado=MG -2.4130071775531983
 Attrib estado=ES 0.28499189425323007
 Attrib estado=SC -1.7006926147150006
 Attrib estado=RS -1.479047186365723
 Attrib estado=RJ -2.4262658345283983
 Attrib estado=PI 0.4036352564811447
 Attrib estado=GO 0.006975382193592268
 Attrib estado=CE 3.1005453129196465
 Attrib estado=SE -0.4086423491101977
 Attrib estado=DF 3.8268622966752153
 Attrib estado=PE -0.8250098631012526
 Attrib estado=BA -1.0814878984367022
 Attrib estado=AM 0.09209296995483959
 Attrib estado=AL -0.39166123164029226
 Attrib estado=MT -0.9401213661516324
 Attrib estado=MS -0.14128189133241617
 Attrib estado=PB 0.3257863141148706
 Attrib estado=TO 0.7192859825853566
 Attrib estado=MA 1.9659380453781639
 Attrib estado=PA 0.2014536221565771
 Attrib estado=RN 0.4059460852219393
 Attrib estado=RO 8.663348595112748
 Attrib estado=AC 0.4959330479809148
 Attrib estado=RR 0.0671323783216531
 Attrib estado=AP 0.22812588560994893
 Attrib estado=Sp 0.18932121856871162
 Attrib estado=sp 0.17160510284358382
 Attrib estado=mg 0.23728563894133187
 Attrib estado=pr 0.15673677651161433
 Attrib estado=Es 0.19730028223821872
 Attrib newsletter -9.540367933148167
 Attrib valorgasto=C 2.4060428865385615
 Attrib valorgasto=B 0.10806483400159171

Attrib valorgasto=A 0.017541481699435268
Attrib pedidos=A -2.246264287217001
Attrib pedidos=B 2.4326270861232393
Attrib pedidos=C 2.3279558795191115
Attrib taxaaprovacao=C 2.373129073834637
Attrib taxaaprovacao=B -2.387515812972836
Attrib taxaaprovacao=A 2.474297251675228
Attrib cadastroprimeiracompra=A 4.469137419732771
Attrib cadastroprimeiracompra=B -1.288659404178382
Attrib cadastroprimeiracompra=C -0.5596934030466579
Attrib primeiraultimacompra=A 4.519094067365142
Attrib primeiraultimacompra=C -0.016587627255122014
Attrib primeiraultimacompra=B -1.9406340765101986

Sigmoid Node 23

Inputs *Weights*

Threshold 0.0765144000488666

Attrib tipopessoa=F -4.213219251228668
Attrib tipopessoa=J -0.687298302991509
Attrib tipopessoa=I 4.823001133400302
Attrib estado=PR -0.22036319344749847
Attrib estado=SP -0.5736822422402148
Attrib estado=MG 0.12053510040075226
Attrib estado=ES -0.16544355756675172
Attrib estado=SC 5.534386685381145
Attrib estado=RS -0.562919661220816
Attrib estado=RJ 0.5764771356693982
Attrib estado=PI 0.38736804743171993
Attrib estado=GO -0.1708895662524845
Attrib estado=CE 0.6683977408658173
Attrib estado=SE 0.496349706204002
Attrib estado=DF 1.0612451286237155
Attrib estado=PE 0.33491103990891374
Attrib estado=BA 0.5458672714793583
Attrib estado=AM 0.4890717384932329
Attrib estado=AL 0.6228881435693864
Attrib estado=MT -1.1667934654864582
Attrib estado=MS 0.4347819471539505
Attrib estado=PB -0.16532350012402
Attrib estado=TO 0.5058530019685387
Attrib estado=MA 0.55841570069682
Attrib estado=PA 0.47853788091333843
Attrib estado=RN 0.5372424761965325
Attrib estado=RO 0.5426643945002285
Attrib estado=AC 0.5233027246408533
Attrib estado=RR 0.4660604372823354
Attrib estado=AP 0.4973103355344124
Attrib estado=Sp 0.49907794349874657
Attrib estado=sp 0.44716844173708775
Attrib estado=mg 0.5357320488741565
Attrib estado=pr 0.4508729096021608
Attrib estado=Es 0.5183427485735154
Attrib newsletter 0.5780174488996689
Attrib valorgasto=C -0.05183755994312399
Attrib valorgasto=B -0.7118103205792953
Attrib valorgasto=A 0.6939572328259128
Attrib pedidos=A 0.21333527620712936
Attrib pedidos=B -0.1765044186388216
Attrib pedidos=C -0.13237874228487187
Attrib taxaaprovacao=C 1.6610865775599797
Attrib taxaaprovacao=B 0.7190964196569443

Attrib taxaaprovacao=A -2.483293752503787
Attrib cadastroprimeiracompra=A -2.145828603643699
Attrib cadastroprimeiracompra=B -1.3376344097862922
Attrib cadastroprimeiracompra=C 3.3887448437279257
Attrib primeiraultimacompra=A 0.983735927881339
Attrib primeiraultimacompra=C -1.8538362437818179
Attrib primeiraultimacompra=B 0.8320541966528086

Sigmoid Node 24

Inputs *Weights*
Threshold 0.18392322845981376
Attrib tipopessoa=F -1.655345809565006
Attrib tipopessoa=J 1.9082729912171492
Attrib tipopessoa=I -0.39485741791800655
Attrib estado=PR -7.861206270991254
Attrib estado=SP -0.12543797605866608
Attrib estado=MG -3.6433568873483804
Attrib estado=ES -3.622398125912691
Attrib estado=SC -2.570105823047163
Attrib estado=RS 2.29610652736348
Attrib estado=RJ 5.131550816318502
Attrib estado=PI -0.25201769293730447
Attrib estado=GO 5.566942087938525
Attrib estado=CE 2.295260980034364
Attrib estado=SE -0.4130088933706905
Attrib estado=DF 3.0426469064790695
Attrib estado=PE -1.1853335667965863
Attrib estado=BA 3.646399065232231
Attrib estado=AM -6.097009303953339
Attrib estado=AL 1.3234373588433272
Attrib estado=MT -2.6165695162533757
Attrib estado=MS 0.9504415582570348
Attrib estado=PB 4.503062336170055
Attrib estado=TO 0.5948222366131786
Attrib estado=MA 3.854485691661731
Attrib estado=PA 0.4187033330693128
Attrib estado=RN 3.851437331659876
Attrib estado=RO 1.3556554645799326
Attrib estado=AC 0.467382994094186
Attrib estado=RR 0.5579486655108932
Attrib estado=AP 0.42893832440385715
Attrib estado=Sp 0.4941260965874867
Attrib estado=sp 0.4577223171213163
Attrib estado=mg 0.455335329717995
Attrib estado=pr 0.46345014076534846
Attrib estado=Es 0.4335216989426703
Attrib newsletter 0.45479679944480084
Attrib valorgasto=C -0.1909014931001337
Attrib valorgasto=B 0.7083558816473804
Attrib valorgasto=A -0.6773753065242963
Attrib pedidos=A 0.06132428654920715
Attrib pedidos=B -0.7513601994060065
Attrib pedidos=C 0.5133536519842933
Attrib taxaaprovacao=C -8.12228534431094
Attrib taxaaprovacao=B 9.70229835902991
Attrib taxaaprovacao=A -1.8048192086884178
Attrib cadastroprimeiracompra=A 4.166922024843008
Attrib cadastroprimeiracompra=B 1.6285785075693473
Attrib cadastroprimeiracompra=C -5.894144811666135
Attrib primeiraultimacompra=A -0.6687622551453495
Attrib primeiraultimacompra=C 3.778146497532959

Attrib primeiraultimacompra=B -3.270441061711638
 Sigmoid Node 25
Inputs Weights
Threshold -1.2225480720754294
Attrib tipopessoa=F -1.7131985912788321
Attrib tipopessoa=J 1.0383775630952619
Attrib tipopessoa=I 1.888776930479849
Attrib estado=PR -1.7135853692441534
Attrib estado=SP -2.8547750481924585
Attrib estado=MG -1.9062832194692707
Attrib estado=ES -0.36639008193034067
Attrib estado=SC 2.766019949631793
Attrib estado=RS 5.382464859375116
Attrib estado=RJ -0.5942760983118529
Attrib estado=PI 0.6152409500633031
Attrib estado=GO -0.005517086077471533
Attrib estado=CE -0.1969284971463719
Attrib estado=SE 0.2423470366234945
Attrib estado=DF -0.6147115275151116
Attrib estado=PE 0.09853747890784292
Attrib estado=BA 0.28589175599501454
Attrib estado=AM 0.3076228333078918
Attrib estado=AL 0.2389261754310385
Attrib estado=MT -0.4493627099424312
Attrib estado=MS 3.6837669890802593
Attrib estado=PB 0.43079810332724605
Attrib estado=TO 0.4214086065923596
Attrib estado=MA 0.2529143142451863
Attrib estado=PA 0.26035411084197024
Attrib estado=RN 0.32552560950598697
Attrib estado=RO 0.3550574413618402
Attrib estado=AC 0.3687049801498488
Attrib estado=RR 0.3193905099201822
Attrib estado=AP 0.3303043784608631
Attrib estado=Sp 0.3019303261321766
Attrib estado=sp 0.30663539160565534
Attrib estado=mg 0.36103535214820187
Attrib estado=pr 0.2894825690203704
Attrib estado=Es 0.2863321609318545
Attrib newsletter -5.715139984046613
Attrib valorgasto=C 1.268024888979298
Attrib valorgasto=B 0.26901094522923885
Attrib valorgasto=A -0.23940478854050398
Attrib pedidos=A -0.7230374490975671
Attrib pedidos=B 1.0649402918951376
Attrib pedidos=C 0.9155499849405169
Attrib taxaaprovacao=C -0.35701943518956814
Attrib taxaaprovacao=B 3.3793290943770096
Attrib taxaaprovacao=A -1.687001420851441
Attrib cadastroprimeiracompra=A 2.6222990933997123
Attrib cadastroprimeiracompra=B -1.957798176670045
Attrib cadastroprimeiracompra=C 0.566268157659312
Attrib primeiraultimacompra=A 1.1600994824330668
Attrib primeiraultimacompra=C -1.5838903533553363
Attrib primeiraultimacompra=B 1.6263721367436288
 Sigmoid Node 26
Inputs Weights
Threshold -0.10909073751784529
Attrib tipopessoa=F -1.6750797054836568
Attrib tipopessoa=J -0.7310066480220512

Attrib tipopessoa=I 2.4874930635355783
Attrib estado=PR 0.18061772393218523
Attrib estado=SP -0.35656979947432654
Attrib estado=MG 0.2492011584320136
Attrib estado=ES 0.2802707951053291
Attrib estado=SC 0.4092075016110195
Attrib estado=RS 0.537487388928503
Attrib estado=RJ 0.3481968870989481
Attrib estado=PI 0.3680580165883383
Attrib estado=GO 0.25667151416514283
Attrib estado=CE 0.4875662217358904
Attrib estado=SE 0.32357118415273156
Attrib estado=DF 0.43116942522476087
Attrib estado=PE 0.32821480425313865
Attrib estado=BA 0.9741513534688864
Attrib estado=AM 0.35348174296553003
Attrib estado=AL 0.42529495115982635
Attrib estado=MT 0.030271754616852827
Attrib estado=MS 0.37744036682572996
Attrib estado=PB 0.41100022471109154
Attrib estado=TO 0.3490301225589385
Attrib estado=MA 0.4829981773382263
Attrib estado=PA 0.3789954670018338
Attrib estado=RN 0.41989572744069603
Attrib estado=RO 0.440140784659121
Attrib estado=AC 0.40015204284091216
Attrib estado=RR 0.3793051642127487
Attrib estado=AP 0.3383378100634005
Attrib estado=Sp 0.35260952587838584
Attrib estado=sp 0.42391826137108174
Attrib estado=mg 0.42035033031863156
Attrib estado=pr 0.3423865260657592
Attrib estado=Es 0.3946785304324871
Attrib newsletter -0.40999299254246935
Attrib valorgasto=C 0.1815330848628211
Attrib valorgasto=B 0.038015625745534645
Attrib valorgasto=A -0.02138990740176264
Attrib pedidos=A -0.04545359581924326
Attrib pedidos=B 0.06724742897143324
Attrib pedidos=C 0.10292025176778151
Attrib taxaaprovacao=C 1.7277860511893024
Attrib taxaaprovacao=B -0.39750423752616104
Attrib taxaaprovacao=A -1.3085096252162178
Attrib cadastroprimeiracompra=A 0.4878194397379125
Attrib cadastroprimeiracompra=B -0.6014184996903821
Attrib cadastroprimeiracompra=C 0.2568612757200345
Attrib primeiraultimacompra=A -0.9925978627129891
Attrib primeiraultimacompra=C 0.5436660139587783
Attrib primeiraultimacompra=B 0.5945462297455371

Sigmoid Node 27

Inputs *Weights*
Threshold -2.167989312438592
Attrib tipopessoa=F -1.1754136141176854
Attrib tipopessoa=J 2.5720889777366174
Attrib tipopessoa=I 0.7538197203039777
Attrib estado=PR -0.768762316822029
Attrib estado=SP -2.584810289786963
Attrib estado=MG 4.368571547186107
Attrib estado=ES -0.7937513971244193
Attrib estado=SC 0.21504254981967444

```

Attrib estado=RS 5.042225429219202
Attrib estado=RJ 1.230902232316378
Attrib estado=PI -0.7621079610005478
Attrib estado=GO -1.9336923075572958
Attrib estado=CE 0.7754766410763948
Attrib estado=SE 0.2709429568179619
Attrib estado=DF 2.4822246423787955
Attrib estado=PE 1.0154626999845748
Attrib estado=BA 1.471511320491157
Attrib estado=AM 0.8072775564199769
Attrib estado=AL -2.739820201112603
Attrib estado=MT 0.9412720627211483
Attrib estado=MS -0.412282952499497
Attrib estado=PB 2.6054908698581905
Attrib estado=TO 0.7174426515352393
Attrib estado=MA -1.6104543043078925
Attrib estado=PA 0.5400252952080993
Attrib estado=RN 1.4890136097425997
Attrib estado=RO 1.013656604159823
Attrib estado=AC 0.8437840421850902
Attrib estado=RR 0.4995287549066775
Attrib estado=AP 0.6371729432045993
Attrib estado=Sp 0.6274239686128097
Attrib estado=sp 0.6486077460238341
Attrib estado=mg 0.6182333253456435
Attrib estado=pr 0.5795390277404199
Attrib estado=Es 0.655825321894424
Attrib newsletter -2.2052565045073553
Attrib valorgasto=C 0.35618323553870174
Attrib valorgasto=B -3.6228957801828385
Attrib valorgasto=A 5.386079808762855
Attrib pedidos=A 3.0546443064961655
Attrib pedidos=B -3.6095180923160863
Attrib pedidos=C 2.639041075099036
Attrib taxaaprovacao=C 11.590081302374221
Attrib taxaaprovacao=B -2.738390576003016
Attrib taxaaprovacao=A -6.7135821191130045
Attrib cadastroprimeiracompra=A 4.397428276281608
Attrib cadastroprimeiracompra=B -2.690356659657739
Attrib cadastroprimeiracompra=C 0.3153805992098023
Attrib primeiraultimacompra=A 2.121787894629507
Attrib primeiraultimacompra=C -0.5720670802081375
Attrib primeiraultimacompra=B 0.5356358838425856

```

Class Sim

Input

Node 0

Class Não

Input

Node 1

Time taken to build model: 1976.41 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	147066	94.4687 %
Incorrectly Classified Instances	8611	5.5313 %
Kappa statistic	0.8542	
Mean absolute error	0.0892	

Root mean squared error	0.2125
Relative absolute error	25.2337 %
Root relative squared error	50.5256 %
Total Number of Instances	155677

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,933	0,017	0,995	0,933	0,963	0,861	0,968	0,991	Sim
	0,983	0,067	0,815	0,983	0,891	0,861	0,968	0,829	Não
Weighted Avg.	0,945	0,029	0,953	0,945	0,946	0,861	0,968	0,954	

=== Confusion Matrix ===

<i>a</i>	<i>b</i>	<-- classified as
111961	7993	<i>a</i> = Sim
618	35105	<i>b</i> = Não